IDENTIFICATION AND CHARACTERIZATION OF CAFFEATE/5 HYDROXYFERULATE 3-O-METHYLTRANSFERASE (COMT) GENES IN *Bambusa tulda*

Project work submitted to the Sister Nivedita University in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

SUBMITTED BY

ADITI KUMARI

Registration No.: 230020281660

Enrollment No.: 2322210002028

DEPARTMENT OF BIOTECHNOLOGY, SISTER NIVEDITA UNIVERSITY

SUPERVISOR

Dr. MALAY DAS

Assistant Professor, Department of Life Sciences

PRESIDENCY UNIVERSITY

School of Health Science and Translational Research

Sister Nivedita University, Kolkata

June 2025

PRESIDENCY UNIVERSITY
KOLKATA
Presidency University
Hindoo college (1817-1855), Presidency college (1855-2010)

Dated-29/05/2025

Supervisor's certificate

This is to certify that **ADITI KUMARI** performed the project work entitled **"IDENTIFICATION AND CHARACTERIZATION OF CAFFEATE/5 HYDROXYFERULATE 3-O-METHYLTRANSFERASE (COMT) GENES FAMILY IN *Bambusa tulda***" under my supervision from January, 2025 to May, 2025.

During the course of the work she has done basic bioinformatics such as database search, BLAST, sequence alignment, phylogenetic tree development, analyses on domain, motifs, promoters, and designing primers. She also performed wet laboratory experiments related to PCR amplification of targeted COMT gene regions in *B. tulda*.

To the best of my knowledge I certify that the data presented in her dissertation was done by her and has not been published anywhere. I wish her every success in all future endeavour.

Malay Das

29/5/2025

Dr. Malay Das
Asst. Professor, Dept. of Life Sciences
Presidency University, Kolkata

86/1, College Street, Kolkata - 700 073, Phone : (033) 2241 1960, Fax : (033) 2241 2738

SISTER NIVEDITA UNIVERSITY
সিস্টার নিবেদিতা বিশ্ববিদ্যালয়

DG Block - 1/2, Action Area I, New Town, Kolkata - 700156, West Bengal, India.

DEPARTMENT OF BIOTECHNOLOGY & DEPARTMENT OF MICROBIOLOGY
জৈব প্রযুক্তি বিভাগ ও অণুজীব বিদ্যা বিভাগ

TO WHOM IT MAY CONCERN

This is to certify that the project report entitled "**IDENTIFICATION AND CHARACTERIZATION OF CAFFEATE/5 HYDROXYFERULATE 3-O-METHYLTRANSFERASE (COMT) GENES IN *Bambusa tulda***", submitted by **Ms. ADITI KUMARI** (**Registration No. 230020281660**, **Enrollment No**. **2322210002028**) to the **Department of Biotechnology**, School of Health Science and Translational Research, Sister Nivedita University, Kolkata is a record of bonafide research works carried out by him/her for the partial fulfilment of the requirements for the Master of Science degree.

Date: 03.06.2025

Place: SISTER NIVEDITA UNIVERSITY

Signature of Head of the Department

Seal

Head
Department of Biotechnology
Sister Nivedita University
Block DG-1/2, Action Area 1
New Town, Kolkata-700 156

A Satyam Roychowdhury Initiative

SNU SISTER NIVEDITA UNIVERSITY

DECLARATION BY THE STUDENT

I hereby declare that the work which is being presented in this dissertation entitled "**IDENTIFICATION AND CHARACTERIZATION OF CAFFEATE/5 HYDROXYFERULATE 3-O-METHYLTRANSFERASE (COMT) GENES IN *Bambusa tulda***" in partial fulfilment of the requirements for the award of the degree of Masters of Science in Biotechnology is an authentic record of my own work carried out during the period from 1st January to 31st May, 2025 under the supervision of **Dr. Malay Das, Assistant Professor, Department of Life Science, Presidency University, Kolkata.**

I acknowledge that I have read and understood the UGC regulations, 2018 (*Promotion of Academic integrity and Prevention of Plagiarism in Higher Educational Institutions*).

I confirm that this dissertation has been checked for plagiarism using the plagiarism checking software **Turnitin**. At the end of the dissertation, a copy of the summary report demonstrating similarities in content and its potential source, generated online using plagiarism checking software is enclosed. I herewith confirm that the dissertation has been less than 10% similarity according to the plagiarism checking software's report and meets the UGC Regulations as well as plagiarism rules of Sister Nivedita University.

I further declare that no portion of the dissertation or its data will be published without the Institute's or supervisor's permission. I have not previously applied for any other degree or award using the topics and findings described in my dissertations.

Aditi Kumari
06/06/25

(Signature of the students with Date)

ADITI KUMARI

Reg No.: 230020281660

Dept.: Biotechnology

ACKNOWLEDGEMENT

I would like to express my gratitude to all associated members of **Sister Nivedita University** and **Presidency University** for giving me the opportunity to complete my dissertation project work successfully. I would like to express my sincere regards to **Dr. Rajat Pal; Assistant Director, Department of Microbiology and Biotechnology** and **Dr. Somsubhra Thakur Choudhury; Head of Department of Biotechnology, Sister Nivedita University, Kolkata** for granting me his consent and approval regarding my Dissertation project at Presidency University.

I render my sincere gratitude to my supervisor **Dr. Malay Das; Assistant Professor of Department of Life Sciences at Presidency University, Kolkata** for considering my credentials, application and for rendering me eligible to carry out my Dissertation project work in his laboratory at Presidency University, for guiding and supervising me, during the entire tenure of my project work. He has enormously encouraged and fulfilled all the requirements to pursue my dissertation work in his laboratory.

I would like to thank my mentor, **Ms. Sudeshna Bera, Junior Research Scholar, Department of Life Sciences, Presidency University**, for her continuous guidance and support throughout the duration of my project. Her unwavering encouragement and dedication ensured that I had all the resources and assistance necessary to carry out my dissertation work successfully under her supervision.

I am also thankful to all Research Scholars **Ms. Mridushree Basak (Senior Research Scholar, Department of Life Sciences, Presidency University), Ms. Sonali Dey (Senior Research Scholar, Department of Life Sciences, Presidency University), Mr. Touhidur Rahman (Senior Research Scholar, Department of Life Sciences, Presidency University), and Ms. Sutrisha Kundu (Junior Research Scholar, Department of Life Sciences, Presidency University)** for their constant guidance, immense support to strength my practical knowledge in the field of plant biotechnology and bioinformatics during the project work.

I would also like to express my profound gratitude to my family and friends for their valuable encouragement and helping throughout this project.

LIST OF FIGURES

Figure No.	Title	Page No.
1	Phenylpropanoid and monolignol biosynthetic pathways	1
2	Plant cell wall consists of Lignin, Cellulose and Hemicellulose (A). Pretreatment causes separation of cellulose and hemicellulose from lignin (B).	5
3	Retrieval of *COMT* genes from various bamboo genome database	11
4	Retrieval of *Bambusa tulda COMT* genes from the GenDBE database	12
5	Construction of phylogenetic tree in MEGA11 software using aligned COMT sequences	12
6	Searching conserved domain using NCBI-CDD	14
7	Graphical conserved domain analysis in TBtools using NCBI-CDD output	14
8	Batch web conserved motif search tool of MEME program	15
9	Clustal Omega tool used for MSA of COMT amino acid sequences	16
10	Promoter analysis in PlantCARE using 2000 base pairs	17
11	Graphical representation of CREs in TBtools	17
12	Primer3 tool layout used for designing gene-specific primers targeting conserved regions of COMT gene sequences in *B. tulda*	18
13	Identification of Btu090.g01022t41(BtCOMT1.1) using GenDBE	22

14	Identification of Btu08A.g01257t29(BtCOMT1.2) using GenDBE	22
15	Identification of rest COMT sequences of *Bambusa tulda* using GenDBE	22
16	Maximum Likelihood (ML) phylogenetic tree constructed using 26 COMT amino acid sequences.	23
17	Maximum Likelihood (ML) phylogenetic tree based on COMT amino acid sequences from species in the BOP clade.	24
18	Graphical representation of conserved domains in COMT proteins, highlighting the structural arrangement of associated domains	26
19	Conserved motif logos of COMT amino acid sequences from *B. tulda* and other bamboo species, identified using the MEME Suite	27
20	Conserved motif distribution among all COMT proteins, where each coloured box represents a specific conserved motif identified by MEME	27
21	Multiple Sequence Alignment of Bamboos with known *COMT* gene of *Hordeum vulgare* were performed and conserved amino acid residues were identified	29
22	Cis-regulatory elements in the 2kb upstream regions of COMT gene in B. tulda. Different colour represents different cis elements	33
23	Gradient PCR showing optimal amplification of *BtCOMT1.1* at 65 °C (left) and *BtCOMT1.2* at 51–52 °C (right) using gene-specific primers.	34

LIST OF TABLES

Table No.	Title	Page No.
1	PCR reaction components and their final concentrations used for amplification in a 50 µL reaction volume	19
2	List of bamboo species with corresponding *COMT* gene family members and their accession numbers retrieved from databases	21
3	Conserved domain prediction table showing the presence of domains in COMT proteins across selected bamboo species	25
4	A series of CREs found in the promoter regions of the *BtCOMT1.1*	31
5	A series of CREs found in the promoter regions of the *BtCOMT1.2*	32
6	Gene-specific primers designed from *COMT* genes in *Bambusa tulda*	34

ABBREVIATION

2G	Second Generation
BLAST	Basic Local Alignment Search Tool
BOP	Bambusoideae, Oryzoideae, and Pooideae
Bp	Base Pair
CDS	Coding DNA Sequence
cDNA	Complementary DNA
gDNA	Genomic DNA
COMT	Caffeate/5-Hydroxyferulate O-Methyltransferase
CRE	Cis-Regulatory Element
DNase I	Deoxyribonuclease I
qRT-PCR	Quantitative Real-Time PCR
MSA	Multiple Sequence Alignment
MEME	Multiple Expectation Maximization for Motif Elicitation
ML	Maximum Likelihood
MEGA	Molecular Evolutionary Genetics Analysis
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction
PLAZA	Plant Genomic Platform
RNA	Ribonucleic Acid
RT-PCR	Real-Time Polymerase Chain Reaction
SAM	S-Adenosyl-L-Methionine
SAH	S-Adenosylhomocysteine
TF	Transcription Factor
TBtools	Toolkit for Biologists
UTR	Untranslated Region

ABSTRACT

The rising global demand for renewable energy has led to increased focus on lignocellulosic biomass as a feedstock for second-generation (2G) bioethanol production. Unlike food-based biofuels, lignocellulosic sources rely on non-edible plant materials, reducing food–fuel competition. However, the high lignin content in plant cell walls restricts enzymatic hydrolysis of cellulose and hemicellulose, making pretreatment both costly and inefficient. As an alternative, genetic modification of lignin biosynthetic pathways offers a promising route to reduce lignin-associated recalcitrance.

Bambusa tulda, a fast-growing tropical bamboo with high cellulose and biomass yield, holds great promise as a bioenergy crop. Despite its industrial potential, the molecular basis of lignin biosynthesis in this species is poorly understood. The caffeate/5-hydroxyferulate O-methyltransferase (COMT) gene family plays a key role in syringyl lignin biosynthesis and is considered a suitable target for lignin engineering.

This study performed a genome-wide identification and characterization of *COMT* genes in *B. tulda*. Multiple sequence alignment and phylogenetic analysis were conducted to examine evolutionary relationships with other monocot species. Functional analyses, including conserved domain and motif prediction, confirmed the structural features of the *COMT* proteins. Promoter regions were analysed to identify cis-regulatory elements associated with stress, light, and hormonal responses.

Furthermore, expression profiling of two functional *COMT* genes across early, middle, and mature tissue stages revealed stage-specific expression patterns linked to lignin deposition. These findings provide the first comprehensive insight into *COMT* gene family members in *Bambusa tulda* and form a foundation for future biotechnological strategies aimed at optimizing lignin content to enhance bioethanol production.

Keywords: Second-generation biofuel, lignocellulosic biomass, lignin biosynthesis, bioethanol, *Bambusa tulda*, *COMT* gene family

CONTENTS

S. No.	TITLE	PAGE NO.
1	INTRODUCTION	1
2	REVIEW OF LITERATURE	4
3	OBJECTIVE	10
4	MATERIALS AND METHODOLOGY	11
5	RESULTS AND DISCUSSION	21
6	CONCLUSION	35
7	FUTURE PROSPECTS	36
8	REFERENCES	37

1. INTRODUCTION

The worldwide shift toward sustainable energy has emphasized the significance of lignocellulosic biofuels as alternatives to fossil fuels, particularly in the context of the rapidly increasing global population and energy demands. Second-generation (2G) bioethanol, derived from lignocellulosic biomass, presents a viable option for decarbonizing transportation sectors without competing with food crops for arable land (Rahaman *et al.*, 2023). However, one of the major challenges in 2G ethanol production is the complex and recalcitrant nature of lignin, an aromatic polymer in plant secondary cell walls, which hinders enzymatic access to cellulose and hemicellulose (Vanholme *et al.*, 2010).

Lignocellulosic biomass comprises three primary components: cellulose, hemicellulose, and lignin. While cellulose and hemicellulose serve as fermentable sugar sources, lignin—a complex aromatic polymer—acts as a physical barrier to enzymatic hydrolysis. Its removal or modification is a major challenge in improving biomass saccharification. Lignin biosynthesis is governed by the phenylpropanoid pathway, which leads to the formation of monolignols such as p-coumaryl, coniferyl, and sinapyl alcohol. These are subsequently polymerized into H (hydroxyphenyl), G (guaiacyl), and S (syringyl) lignin units (Fig.1), respectively (Boerjan *et al.*, 2003).

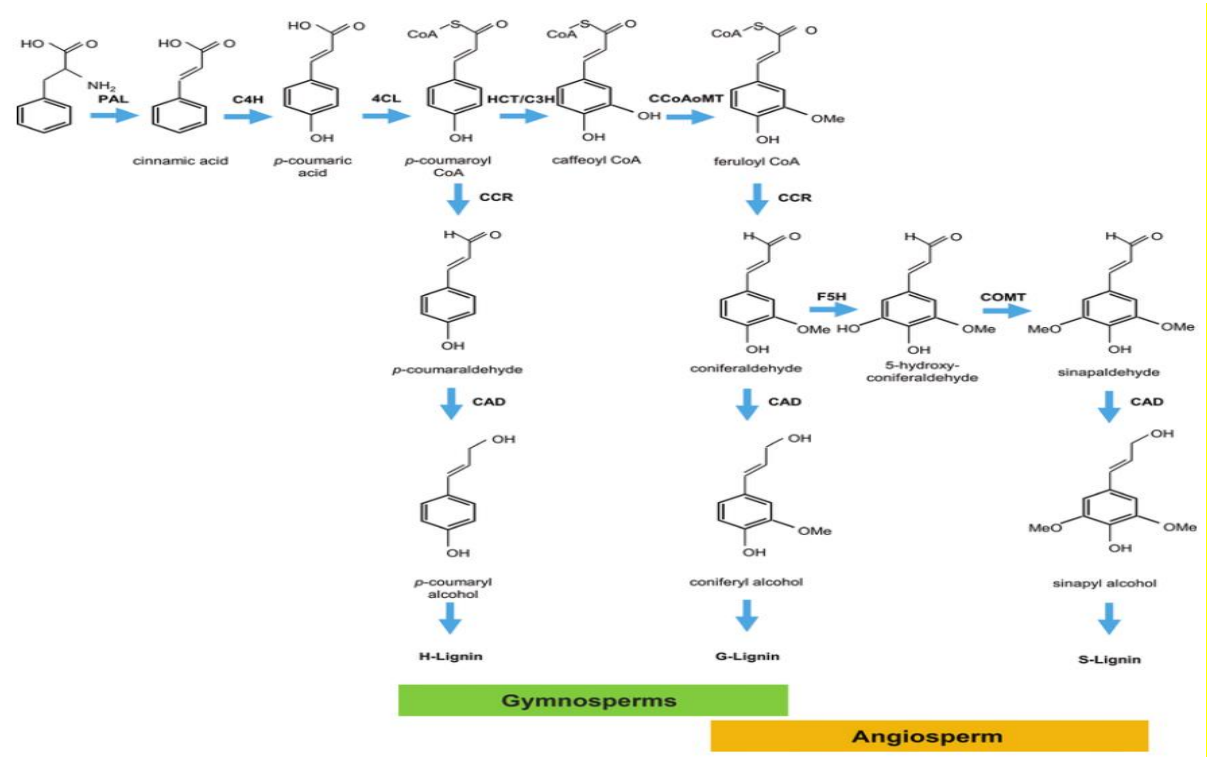


Fig.1: Phenylpropanoid and monolignol biosynthetic pathways

One of the key enzymes in this pathway is caffeate/5-hydroxyferulate O-methyltransferase (COMT), which catalyzes the methylation of 5-hydroxyferulic acid into sinapic acid, a direct precursor of syringyl lignin. COMT plays a crucial role in determining the S/G ratio of lignin, which influences both lignin solubility and biomass digestibility. Studies in model and crop species—including *Arabidopsis thaliana*, *Oryza sativa*, *Hordeum vulgare* and *Zea mays*—have demonstrated that downregulation or knockout of COMT leads to reduced lignin content, altered S/G ratios, and improved sugar release during enzymatic hydrolysis (Halpin *et al*., 1998; Vanholme *et al*., 2012; Zhang *et al*., 2019).

A comprehensive understanding of *COMT* gene family members in *B. tulda* requires a genome-wide identification approach based on comparative genomics. Grasses belonging to the BOP (Bambusoideae-Oryzoideae-Pooideae) clades provide a rich dataset for phylogenetic comparison, as functional COMT orthologs have already been validated in species like *Brachypodium distachyon* (BdCOMT), *Oryza sativa* (OsCOMT), *Phyllostachy edulis* (PeCOMT) (Shen *et al.*, 2013; Meng *et al.*, 2021). Phylogenetic analysis of COMT homologs across these species can help identify conserved and divergent evolutionary patterns relevant to function.

Targeting COMT through genetic engineering is an emerging strategy to reduce lignin recalcitrance and enhance biomass conversion efficiency. The manipulation of COMT expression, either through RNA interference, gene editing, or promoter modification, offers a cost-effective and sustainable alternative to energy- and chemical-intensive pretreatment methods. Functional genomics studies have further revealed that *COMT* genes possess highly conserved domains—such as Methyltransf_2 (PF00891)—and regulatory elements in their promoter regions that respond to abiotic stress, hormones, and developmental signals (Zhao *et al.*, 2013; Wang *et al.*, 2022).

Despite considerable advances in lignin pathway research across various monocots and dicots, bamboo remains underexplored. Bamboo, a fast-growing perennial grass, has emerged as a promising lignocellulosic feedstock for 2G bioethanol production. With its rapid biomass accumulation, high cellulose content, and adaptability to marginal lands, bamboo fits well within the sustainability criteria for bioenergy crops (Scurlock *et al.*, 2000). Among Indian bamboo species, *Bambusa tulda* stands out due to its commercial importance and favorable biochemical profile. (Biswas *et al.*, 2022). A previous study demonstrated that *B. tulda* had a

high α-cellulose content (~65%) and moderate lignin levels (~20–23%), making it suitable for bioethanol production.

However, the genetic architecture of lignin biosynthesis in *B. tulda* remains largely unexplored. While *COMT* gene families have been well-characterized in species such as *Brachypodium distachyon*, *Oryza sativa*, and *Phyllostachys edulis*, little is known about their homologs in bamboo. Comparative genomics and phylogenetic studies using these reference species can aid in identifying orthologous *COMT* genes in *B. tulda* and understanding their evolutionary and functional divergence (Shen *et al.*, 2013; Meng *et al.*, 2021).

This research focuses on the genome-wide identification and characterization of *COMT* genes in *Bambusa tulda*. By employing homology-based searches, domain architecture analysis, motif discovery, promoter cis-element prediction, and phylogenetic reconstruction, this study aims to elucidate the structural and regulatory features of *COMT* genes. Such insights will be instrumental for future efforts to engineer *B. tulda* varieties with optimized lignin profiles for enhanced saccharification and biofuel yield.

Ultimately, understanding the genetic basis of lignin biosynthesis in bamboo can contribute to sustainable bioenergy solutions, offering a low-cost, non-food lignocellulosic resource for 2G bioethanol production. The outcomes of this study will also serve as a foundation for advanced biotechnological interventions in bamboo and related grass species.

2. REVIEW OF LITERATURE

2.1 Bioenergy and the Evolution of Feedstock Generations:

The global demand for renewable energy has surged in response to escalating concerns over climate change, fossil fuel depletion, and greenhouse gas emissions. Among renewable sources, bioenergy—particularly in the form of biofuels—has emerged as a vital component of sustainable energy systems. Biofuels are broadly classified into first-generation (1G) and second-generation (2G) categories based on the nature of the biomass feedstock used and the technological processes involved.

First-generation biofuels are produced from food crops such as corn, sugarcane, and soybean by fermenting sugars or extracting oils. Although these biofuels have achieved commercial viability, their widespread use raises concerns about food security, land-use change, and environmental sustainability (Naik *et al.*, 2010). Additionally, the energy balance and greenhouse gas savings associated with 1G biofuels are often marginal when indirect land use and agricultural emissions are accounted for.

Second-generation biofuels, on the other hand, utilize lignocellulosic biomass such as crop residues, forestry waste, and dedicated energy crops like switchgrass, miscanthus, and bamboo. These feedstocks are non-edible, widely available, and less dependent on agricultural inputs, making them a more sustainable alternative. However, the conversion of lignocellulosic biomass into fermentable sugars is technologically challenging due to the complex structure of the plant cell wall, particularly the presence of lignin—a phenolic polymer that provides rigidity and resistance to microbial degradation (Zhao *et al.*, 2012).

Advanced research and innovation in biotechnology, genomics, and systems biology are now playing critical roles in optimizing biomass crops and improving conversion technologies to realize the full potential of 2G biofuels globally.

2.2 Importance of Pretreatment in Biomass Conversion:

Pretreatment is a crucial upstream process in the biochemical conversion of lignocellulosic biomass, aimed at altering the physical and chemical structure of the biomass to make cellulose and hemicellulose more accessible to hydrolytic enzymes (Fig.2). Effective

pretreatment not only increases sugar yields but also reduces enzyme requirements and improves the overall economics of biofuel production.

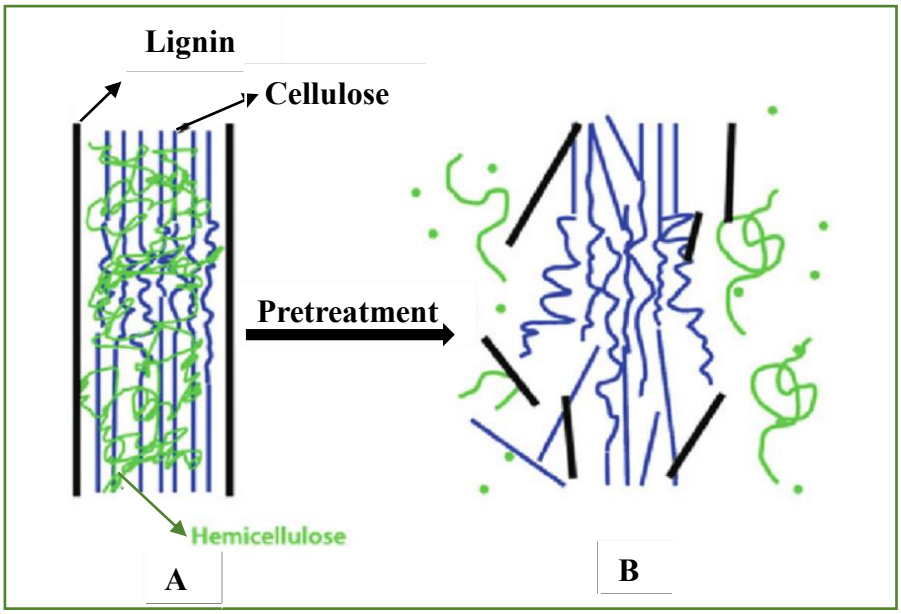


Fig.2: Plant cell wall consists of Lignin, Cellulose and Hemicellulose (A). Pretreatment causes separation of cellulose and hemicellulose from lignin (B)

Various pretreatment methods have been developed over the years. In physical pretreatment such as milling, grinding and ultra-sonication are used to increase accessibility of surface area, decrease particle size, decrease polymerization degree, and increase efficiency of enzymatic hydrolysis, while chemical pretreatment methods such as dilute acids, dilute alkalis, ionic liquids, organic solvents, etc. are used. When acid is used for pretreatment, it dissolves hemicellulose and modifies lignin. Acid pretreatment increases the exposure of cellulose to hydrolytic enzymes. When basic treatment is given, it breaks the ester linkage between cellulose and lignin and increases the exposure of cellulose to hydrolytic enzymes. Again, alkali catalysts are expensive, require the neutralization of surplus alkali, and sometimes take a long time. Hydrophobic pretreatments are done for amphiphilic molecules in order to dissolve soluble organic compounds in aqueous solutions. In the physico-chemical pretreatment method, steam explosion, hot water pretreatment, and supercritical carbon dioxide explosion are done. But various inhibitory products like furfural, alcohols, aldehydes, etc. are produced. Hot water pretreatment is used for hydrolysis of hemicellulose in order to obtain pure hemicellulose, but again, the process demands a huge amount of energy. Organic solvent pretreatment is used for the prehydrolysis of hemicellulose and breaking of bonds between lignin and hemicellulose (Haldar *et al.*, 2021). Ozonolysis is also used which causes

fractional degradation of hemicellulose without inhibiting production of end products. Biological pretreatment methods are also there which uses microorganisms for delignification of lignocellulosic masses like *Phelbia radiata*, *Dichmitus squalens. Rigidosporous lignisux*, etc. Using biological methods of pretreatment produces less hazardous wastes. But again maintaining optimum growth conditions in the bioreactors are not easy to maintain and also it is more expensive.

In the context of bamboo biomass, pretreatment plays a pivotal role in overcoming the challenges posed by its high lignin content and dense culm structure. Alkaline and oxidative methods have been particularly effective in delignifying bamboo without extensive carbohydrate loss. For example, in a study involving Indian bamboo species, including *Bambusa tulda*, hydrothermal pretreatment improved cellulose accessibility and enzymatic hydrolysis (Biswas *et al.,* 2022). Nonetheless, pretreatment processes often involve high energy input, production of inhibitory by-products (e.g., furfurals and phenolics), and operational complexity, underscoring the need for genetic-level interventions to complement or reduce reliance on such methods.

2.3 Rationale for Genetic Modification over Pretreatment:

While pretreatment remains a vital component of lignocellulosic bioethanol production, its limitations—such as high cost, incomplete lignin removal, and generation of toxic inhibitors—highlight the need for upstream genetic interventions. Modifying the lignin biosynthesis pathway through genetic engineering offers a promising solution to these bottlenecks. Specifically, targeting genes involved in monolignol biosynthesis, such as COMT, enables the development of plant varieties with reduced lignin content or altered lignin composition, thereby enhancing cell wall digestibility.

Several studies have demonstrated the feasibility of this approach. In *Arabidopsis thaliana*, downregulation of COMT led to a reduction in syringyl lignin and an improvement in biomass saccharification (Vanholme *et al.,* 2012). In maize, the well-known *bm3* mutant, which carries a defective *COMT* gene, exhibited improved digestibility and higher ethanol yields (Halpin *et al.,* 1998). Similarly, in switchgrass, RNAi-mediated silencing of COMT resulted in significant reductions in lignin and improvements in sugar release (Fu *et al.,* 2011).

Unlike pretreatment, which acts post-harvest, genetic modification operates at the source by engineering the plant itself for improved processability. This approach has the potential to reduce pretreatment severity, enzyme load, and overall processing cost, while maintaining or even improving biomass yield. Thus, the integration of genetic engineering into biofuel crop development is becoming a cornerstone of next-generation biorefinery strategies.

2.4 Lignin Biosynthesis and the Role of COMT:

Lignin is a major structural polymer in vascular plants, comprising 15–30% of the total dry biomass. It provides mechanical strength, water transport efficiency, and pathogen resistance. However, its cross-linked aromatic structure poses a significant barrier to enzymatic digestion of cellulose and hemicellulose. Understanding and manipulating the lignin biosynthesis pathway is therefore central to improving biomass deconstruction.

The phenylpropanoid pathway begins with phenylalanine, which is converted into cinnamic acid by phenylalanine ammonia lyase (PAL). Through a series of enzymatic steps involving cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT), and other enzymes, monolignol precursors are formed. COMT catalyzes the methylation of 5-hydroxyferulic acid to sinapic acid, a key precursor for syringyl lignin synthesis (Boerjan *et al.*, 2003).

COMT activity influences the syringyl-to-guaiacyl (S/G) ratio, which affects lignin's structural properties and susceptibility to degradation. Mutants with impaired COMT function often exhibit a shift in lignin composition, leading to more amenable biomass for enzymatic hydrolysis. Recent advancements in transcriptomics and proteomics have allowed deeper insights into COMT gene expression, regulation, and interaction with other lignin pathway genes, enabling more precise genetic interventions (Mottiar *et al.*, 2016).

2.5 Bamboo as a Lignocellulosic Bioenergy Crop:

Bamboo belongs to the Poaceae family and is one of the fastest-growing plants on Earth, with some species growing up to a meter per day. Its high productivity, short harvest cycle (3–5 years), and ability to grow on marginal lands make it an attractive bioenergy feedstock. Moreover, bamboo has a high lignocellulosic content, with cellulose and hemicellulose

making up about 60–70% of the biomass and lignin accounting for 15–25% depending on species and age.

Bambusa tulda, in particular, is widely distributed in South and Southeast Asia and is used extensively in construction and handicrafts. It has recently attracted attention for bioethanol production. Biswas et al. (2022) conducted a comparative study on seven Indian bamboo species and found that *B. tulda* exhibited favourable properties such as moderate lignin content (~20–23%), high α-cellulose content (~65%), and good thermal stability. Anatomically, *B. tulda* features well-developed vascular bundles and thick-walled fibers that contribute to its mechanical strength while still allowing for effective enzymatic digestion post-treatment.

Additional studies have indicated that the chemical composition and enzymatic digestibility of bamboo can vary significantly with age, culm height, and environmental factors (Scurlock *et al.*, 2000). This variability underscores the need for genetic-level understanding and improvement to develop uniform and process-optimized bamboo varieties for bioenergy applications.

2.6 Identification and Characterization of *COMT* Genes in *Bambusa tulda:*

With the availability of bamboo genomic data through genome databases like the BambooBase, Monocots PLAZA 5.0, and GenDBE, it has become possible to identify and characterize gene families involve in lignin synthesis. In this study, *COMT* genes from *Hordeum vulgare* were used as queries to identify homologous sequences in *Bambusa tulda* through BLAST analysis.

The retrieved *COMT* gene sequences were validated using conserved domain analysis tools such as NCBI-CDD, Pfam, and TBtools. Functional domains like Methyltransf_2 (PF00891) and PF08100, essential for S-adenosyl methionine binding and catalytic activity, were identified. Clustal Omega was used for multiple sequence alignment, confirming conserved motifs across species. MEME Suite identified 12 conserved motifs, which were mapped to functional regions of the protein. Phylogenetic analysis using MEGA11 grouped *B. tulda* COMTs with homologs from BOP Clades, indicating evolutionary conservation and potential orthology.

These findings not only confirm the presence of functional COMT homologs in *B. tulda* but also provide a molecular basis for further functional characterization and gene editing strategies.

2.7 Functional Genomics and Future Perspectives:

Functional genomics has played a pivotal role in uncovering the biological roles of lignin biosynthetic genes, including COMT. Experimental models like maize, rice, and switchgrass have demonstrated that manipulating COMT expression can lead to significant improvements in lignocellulosic biomass digestibility. CRISPR/Cas9 gene editing, in particular, has revolutionized the ability to make precise, heritable modifications in target genes. In rice, CRISPR-induced disruption of the *OsCOMT* gene improved saccharification efficiency and ethanol yield without compromising plant development (Zhang *et al.*, 2019).

In the context of *Bambusa tulda*, functional studies remain limited. However, the identification of *COMT* gene family members and their conserved domains provides a strong foundation for future research. Tissue-specific expression analysis, promoter characterization, and gene knockout or overexpression studies will be instrumental in validating *COMT* gene functions in bamboo. Additionally, integrating omics technologies—such as transcriptomics, proteomics, and metabolomics—can facilitate the discovery of regulatory networks and co-expressed gene clusters associated with lignin metabolism.

Such integrative approaches will accelerate the development of genetically engineered bamboo varieties tailored for high-efficiency bioethanol production and support the broader goals of sustainable bioenergy and climate resilience.

3. OBJECTIVE

In lignin biosynthesis pathway there are many genes involved but among them there are two genes COMT and F5H are mainly involved in formation of syringyl or S unit of the lignin. As mentioned in previous discussion our main aim is to decrease the S unit formation of lignin to increase biomass of the plants which leads to better saccharification yield and better glucose amount for fermentation of bioethanol. Understanding the functional *COMT* gene and identifying it in phylogenetically related grasses will help in future to target on these genes in various bamboos by performing the following analyses:

- Database search and sequences retrieval of all *COMT* gene in Bamboos (*Ampelocalamus luodianensis, Bonia amplexicaulis, Raddia guianensis, Hsuehochloa calcarean, Dendrocalamus sinicus, Guadua angustifolia, Melocanna baccifera, Olyra latifolia, Otatea glauca, Phyllostachys edulis, Raddia guianensis, Rhipidocladum racemiflorum, Bambusa tulda)*
- Phylogenetic analysis of all *COMT* gene in Bamboos
- Analysis on conserved domain of all *COMT* gene in Bamboos
- Analysis on conserved motif of all *COMT* gene in Bamboos
- Analysis on conserved amino acid residues of all *COMT* gene in Bamboos
- Analysis on promoter and conserved Cis regulatory elements of target *COMT* genes in *Bambusa tulda*
- Expression analysis of target *COMT* genes in stem tissues at different developmental stages of *Bambusa tulda* by designing gene specific primers from above studies

4. MATERIALS AND METHODS

4.1. Database search and retrieval of homologous COMT sequences from bamboo species (*Ampelocalamus luodianensis, Bonia amplexicaulis, Raddia guianensis, Hsuehochloa calcarean, Dendrocalamus sinicus, Guadua angustifolia, Melocanna baccifera, Olyra latifolia, Otatea glauca, Phyllostachys edulis, Raddia guianensis, Rhipidocladum racemiflorum, Bambusa tulda)*:

To identify lignin biosynthesis *COMT* gene family in Bamboos, *COMT* gene sequence, previously characterized to function in lignin biosynthesis from *Hordeum vulgare* (Jae *et al.,* 2020) was used as query sequence in a BLAST search tool in BambooBase (https://bamboo.genobank.org/) (Fig.3) and in GenDBE (https://gendbe.computational.bio.uni-giessen.de/cgi-bin/login.cgi?cookie_test=1) (Fig.4).

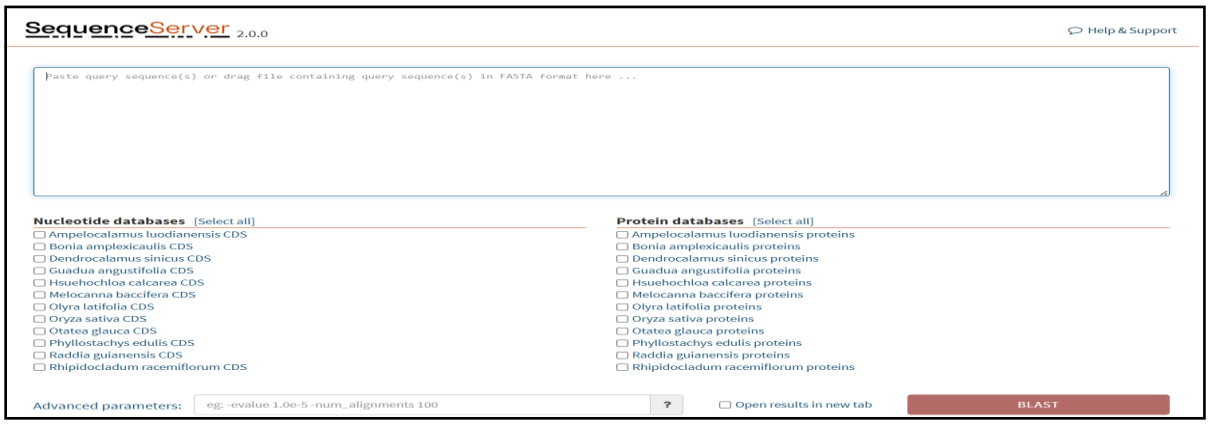


Fig.3: All *COMT* genes from various bamboo species were retrieved from the database

In BambooBase, all the species of bamboos was selected and "Search by BLAST" was performed after entering the *HvCOMT* sequence which were already characterized as functional *COMT* gene.

- All the COMT sequences which were chosen for study were obtained from *Ampelocalamus luodianensis, Bonia amplexicaulis, Raddia guianensis, Hsuehochloa calcarean, Dendrocalamus sinicus, Guadua angustifolia, Melocanna baccifera, Olyra latifolia, Otatea glauca, Phyllostachys edulis, Rhipidocladum racemiflorum*. All the sequences which were obtained from BLAST-P search tool of BambooBase were with e-value = 0 with 100% query coverage.

Subsequently, all *COMT* amino acid sequences CDS of *Bambusa tulda* were retrieved from GenDBE.

- And all the retrieved *COMT* gene sequences of *Bambusa tulda* were further analysed using the same platform for their gene structure and functional analysis.

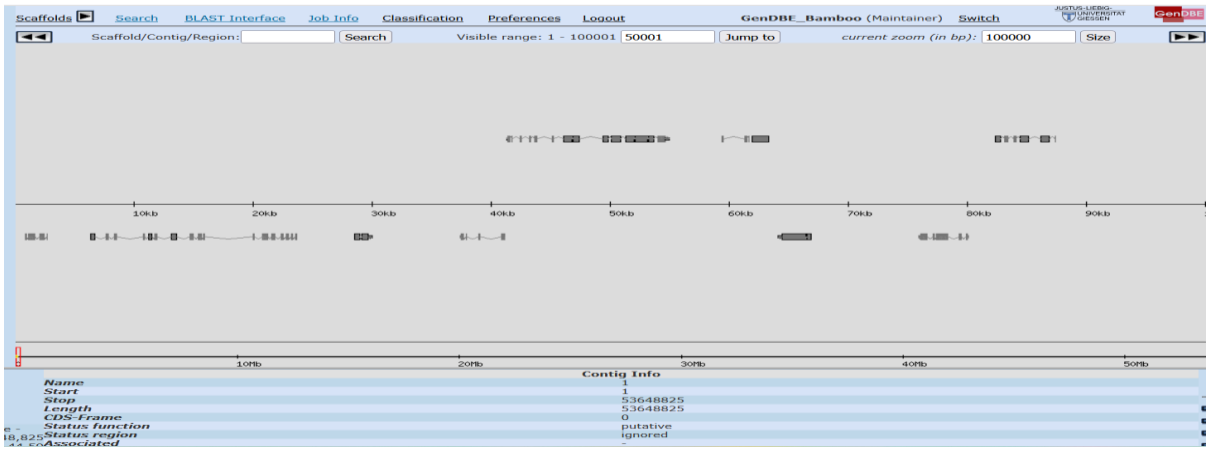


Fig.4: All *COMT* genes from *Bambusa tulda* were retrieved from the GenDBE

4.2. Phylogenetic analysis of *COMT* gene in Bamboos:

All the retrieved COMT sequences from both BambooBase (*Ampelocalamus luodianensis, Bonia amplexicaulis, Raddia guianensis, Hsuehochloa calcarean, Dendrocalamus sinicus, Guadua angustifolia, Melocanna baccifera, Olyra latifolia, Otatea glauca, Phyllostachys edulis, Rhipidocladum racemiflorum*) and GenDBE (*Bambusa tulda*) were subjected to multiple sequence alignment, and a Maximum Likelihood (ML) unrooted phylogenetic tree was constructed using MEGA11 software (Fig.5) by using default parameters to show similarities or differential relationships among all the Bamboos.

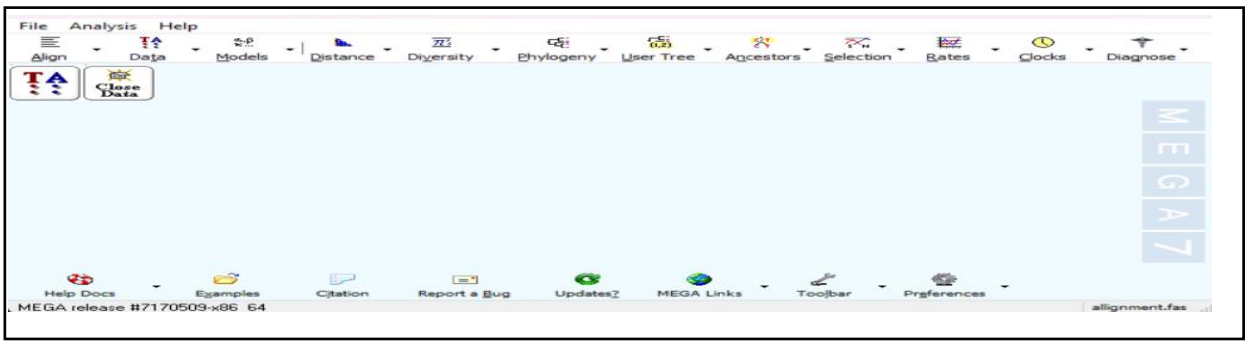


Fig.5: The phylogenetic tree was constructed in MEGA11 software using aligned COMT sequences

For further analysis, COMT sequences from all the BOP clade members were retrieved from Monocots PLAZA5.0(https://bioinformatics.psb.ugent.be/plaza.dev/instances/monocots_05/). A phylogenetic tree was constructed using all the COMT sequences with similar conditions as mentioned above.

- The multiple sequence alignment (MSA) file was also prepared by using MEGA11 software before the tree analysis.

4.3. Conserved signature domain analysis of *COMT* gene in Bamboos:

Conserved domains of all retrieved COMT sequences from bamboo species were analysed using the NCBI Conserved Domain Database (CDD) (Fig.6) through the CD-Search tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). The resulting domain data were analysed and visualized using TBtools II (Toolbox for Biologists) version 2.119 (https://github.com/CJ-Chen/TBtools).

- All the retrieved amino acid sequences of the *COMT* family gene file were uploaded and with an E-mail, it was submitted after that the following page will show up where Click the 'Browse result' tab. A new page will open. On the left side, under "Select query sequences" select all the IDs, select "Compact mode" and press "Show selected queries", Where the following image of domains will open.

The conserved domain file (hitdata file) downloaded from the NCBI Conserved Domain Database (NCBI-CDD), containing conserved domain information for all retrieved COMT amino acid sequences, was uploaded to TBtools (Fig.7) for further analysis.

- The graphical output generated by TBtools was used to examine domain architecture, and these conserved domains served as key markers for verifying and validating the functional identity of each *COMT* gene.

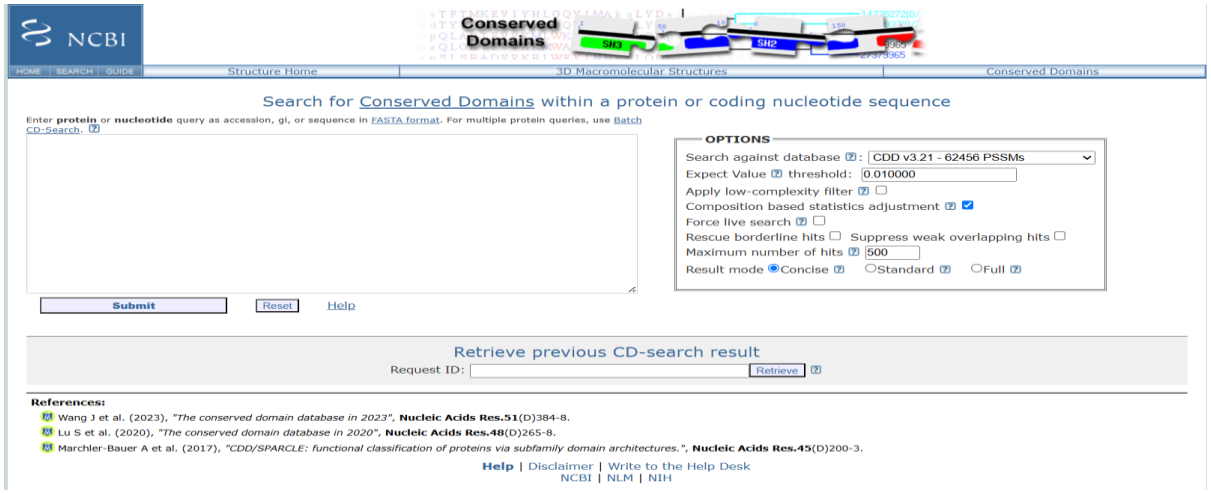

NCBI
Conserved Domains
Structure Home
3D Macromolecular Structures
Conserved Domains
Search for Conserved Domains within a protein or coding nucleotide sequence
OPTIONS
Search against database: CDD v3.21 - 62456 PSSMs
Expect Value threshold: 0.010000
Apply low-complexity filter
Composition based statistics adjustment
Force live search
Rescue borderline hits
Suppress weak overlapping hits
Maximum number of hits: 500
Result mode Concise Standard Full
Submit
Reset
Help
Retrieve previous CD-search result
Request ID:
Retrieve
References:
Wang J et al. (2023), "The conserved domain database in 2023", Nucleic Acids Res.51(D)384-8.
Lu S et al. (2020), "The conserved domain database in 2020", Nucleic Acids Res.48(D)265-8.
Marchler-Bauer A et al. (2017), "CDD/SPARCLE: functional classification of proteins via subfamily domain architectures.", Nucleic Acids Res.45(D)200-3.
Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

Fig.6: Batch web conserved domain search tool of NCBI-CDD

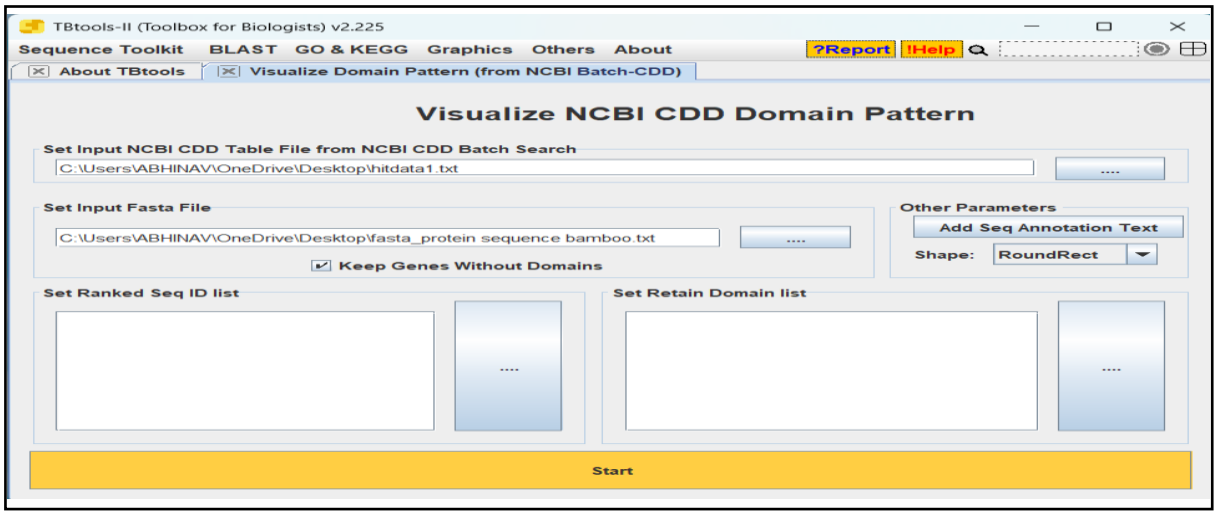


Fig.7: Graphical conserved domain analysis in TBtools using NCBI-CDD output

4.4. Conserved motif analysis of *COMT* sequences in Bamboos:

To examine differences among all the retrieved COMT amino acid sequences, conserved motifs were identified using the MEME Suite (Multiple Expectation Maximization for Motif Elicitation) available at https://meme-suite.org/meme/tools/meme. The analysis was performed with the parameters set to a maximum of 12 motifs, with motif widths ranging from 6 to 50 amino acids (Fig.8).

- In MEME program, under the "Upload Sequences" tab, the FASTA file containing the all the retrieved COMT amino acid sequences were uploaded. Under "How many

motifs should MEME find?", 12 motifs are entered. Under Advanced options, the Minimum width is kept 6 while the Maximum width is kept 50. All other settings are kept as it is. The 'Start Search' option is selected.

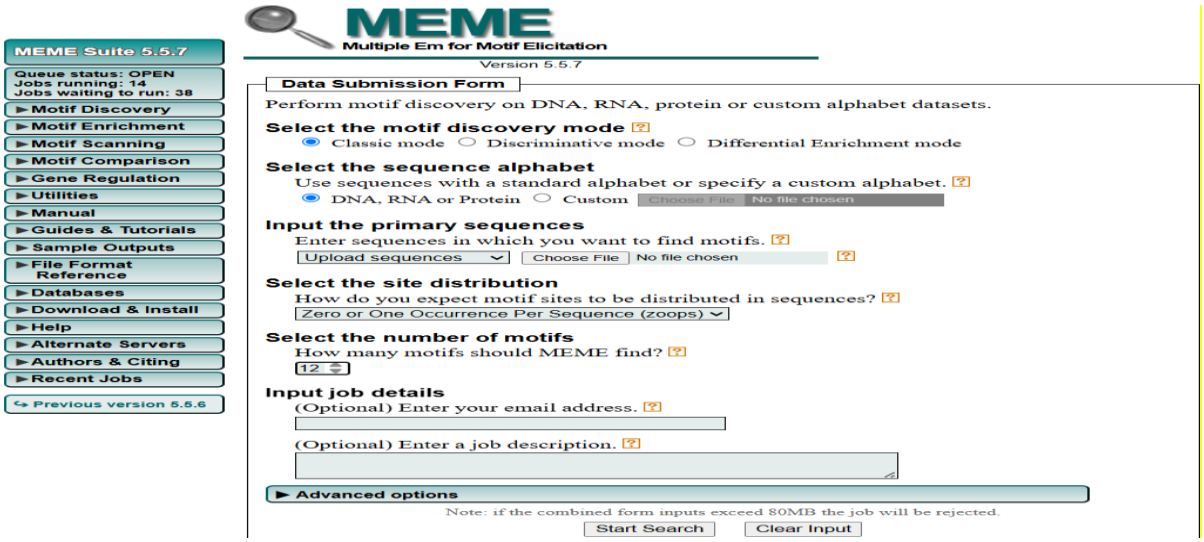


Fig.8: Batch web conserved motif search tool of MEME program

4.5. Analysis of conserved amino acid residues in *COMT* sequences of Bamboos:

To identify conserved amino acid residues in the COMT gene family of bamboos, multiple sequence alignment (MSA) was performed using all retrieved COMT protein sequences. The alignment was carried out using the Clustal Omega server (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) to assess sequence similarity and detect conserved regions across species (Fig.9).

- Based on the aligned sequences, conserved amino acid residues were identified by comparing bamboo COMT proteins with known COMT sequences from other plant species, supporting the confirmation of COMT gene identity and functional conservation.

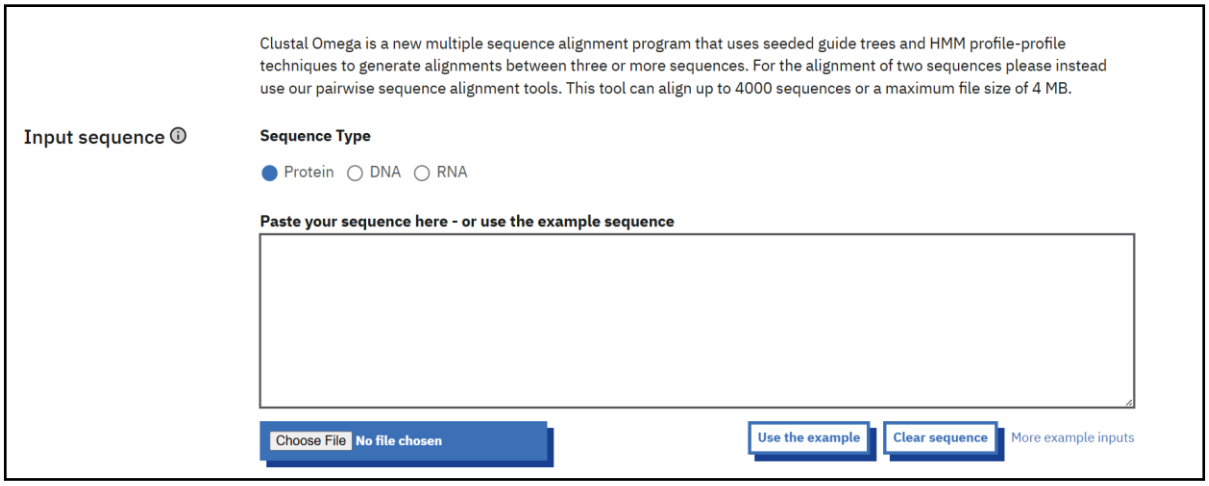

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools. This tool can align up to 4000 sequences or a maximum file size of 4 MB.
Input sequence
Sequence Type
Protein
DNA
RNA
Paste your sequence here - or use the example sequence
Choose File
No file chosen
Use the example
Clear sequence
More example inputs

Fig.9: Clustal Omega tool used for MSA of COMT amino acid sequences

4.6. Promoter and conserved Cis regulatory element analysis of target *COMT* genes in *Bambusa tulda*:

The 2000 base pair upstream regions from the transcription start site of the functional *COMT* genes were extracted from GenDBE and the cis-regulatory elements from extracted promoter sequences of *BtCOMT* were analysed using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) (Fig.10).

- In PlantCARE server, under "sequence to submit" the 2Kb upstream sequence of each COMT gene was pasted in FASTA format. In "email address to send the result back", e-mail ID was submitted. The Search option is selected. The result can be downloaded from the mail.

- Promoters were submitted in FASTA format to these tools, and the annotated motifs were classified based on their associated biological function (e.g., development response, stress response, light response, etc.).

The file containing information regarding cis regulatory elements were downloaded for further analysis.

- The graphical output of desired CREs was generated by TBtools (Fig.11) for further representation.

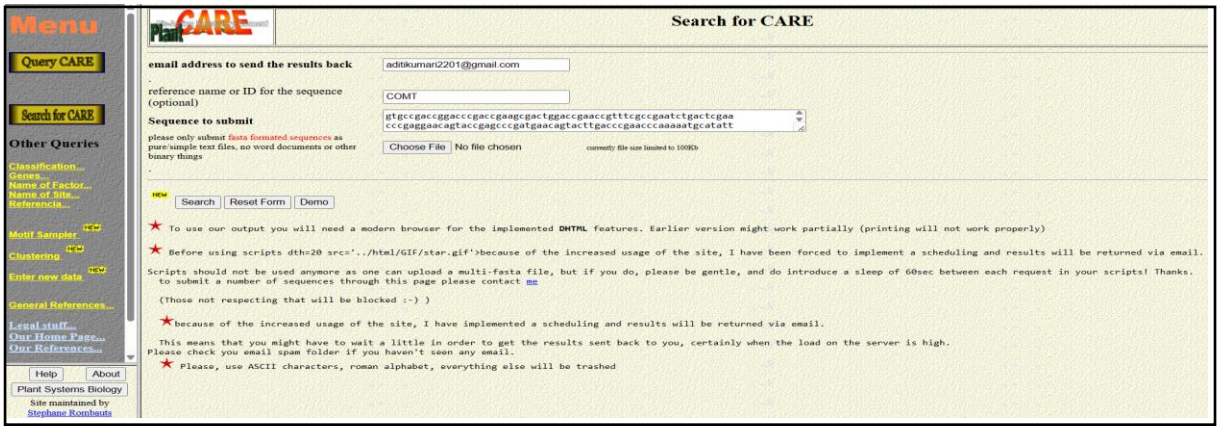

Menu
Query CARE
Search for CARE
Other Queries
Classification...
Genes...
Name of Factor...
Name of Site...
Referencia...
Motif Sampler
Clustering
Enter new data
General References...
Legal stuff...
Our Home Page...
Our References...
Help
About
Plant Systems Biology
Site maintained by
Stephane Rombauts
Search for CARE
email address to send the results back
aditikumari2201@gmail.com
reference name or ID for the sequence (optional)
COMT
Sequence to submit
Choose File
No file chosen
Search
Reset Form
Demo
To use our output you will need a modern browser for the implemented DHTML features. Earlier version might work partially (printing will not work properly)
Before using scripts dth=20 src='../html/GIF/star.gif'>because of the increased usage of the site, I have been forced to implement a scheduling and results will be returned via email.
Scripts should not be used anymore as one can upload a multi-fasta file, but if you do, please be gentle, and do introduce a sleep of 60sec between each request in your scripts! Thanks.
to submit a number of sequences through this page please contact me
(Those not respecting that will be blocked :-))
because of the increased usage of the site, I have implemented a scheduling and results will be returned via email.
This means that you might have to wait a little in order to get the results sent back to you, certainly when the load on the server is high.
Please check you email spam folder if you haven't seen any email.
Please, use ASCII characters, roman alphabet, everything else will be trashed

Fig.10: Promoter analysis in PlantCARE using 2000 base pairs

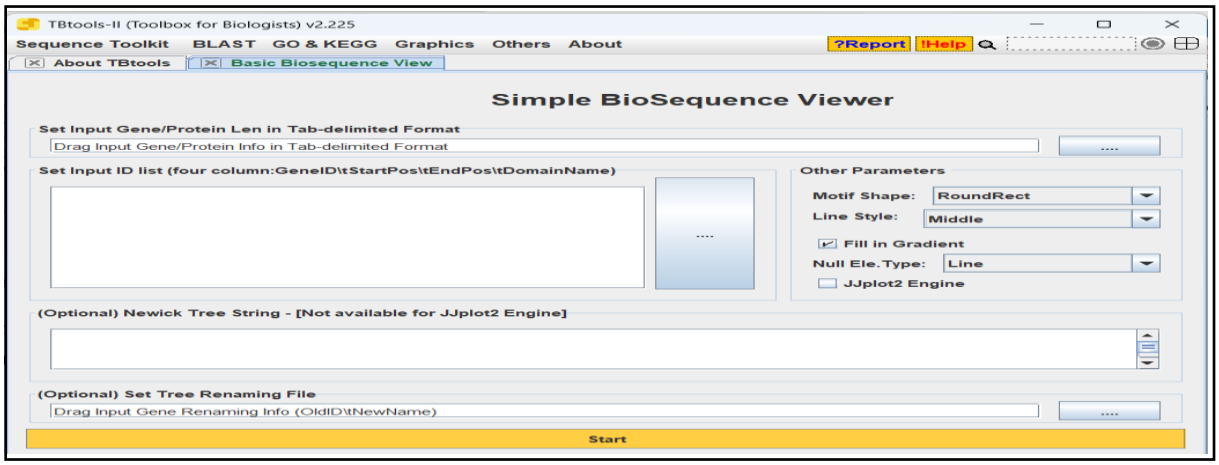


Fig.11: Graphical representation of CREs in TBtools

4.7. Expression Analysis of *COMT* genes in Stem Tissue of *Bambusa tulda* using Real-Time PCR:

4.7.1. Isolation of total RNA and preparation of cDNA libraries:

"cDNA for real-time PCR analysis from selected tissues at varying developmental stages (early, middle, and mature) was previously prepared in the lab."

- Total RNA was extracted from selected *Bambusa tulda* tissues representing early, middle, and mature developmental stages using TRIzol reagent and the RNeasy Plant Mini Kit (Qiagen). To eliminate potential genomic DNA contamination, RNA samples

were treated with DNase I (Dutta *et al.*, 2021). The concentration and purity of the isolated RNA were assessed using a Spectrophotometer, and RNA integrity was verified by 1% (w/v) agarose gel electrophoresis with formamide. Approximately 1 µg of total RNA from each sample was used for first-strand cDNA synthesis using the Verso cDNA Synthesis Kit (Thermo Fisher Scientific), following the manufacturer's instructions (Chakraborty *et al.,* 2019).

4.7.2. Primer designing:

The CDS sequences of functional COMT genes in *Bambusa tulda* were aligned using Clustal Omega, and primers were designed using Primer3 software (Fig.12) by targeting conserved regions.

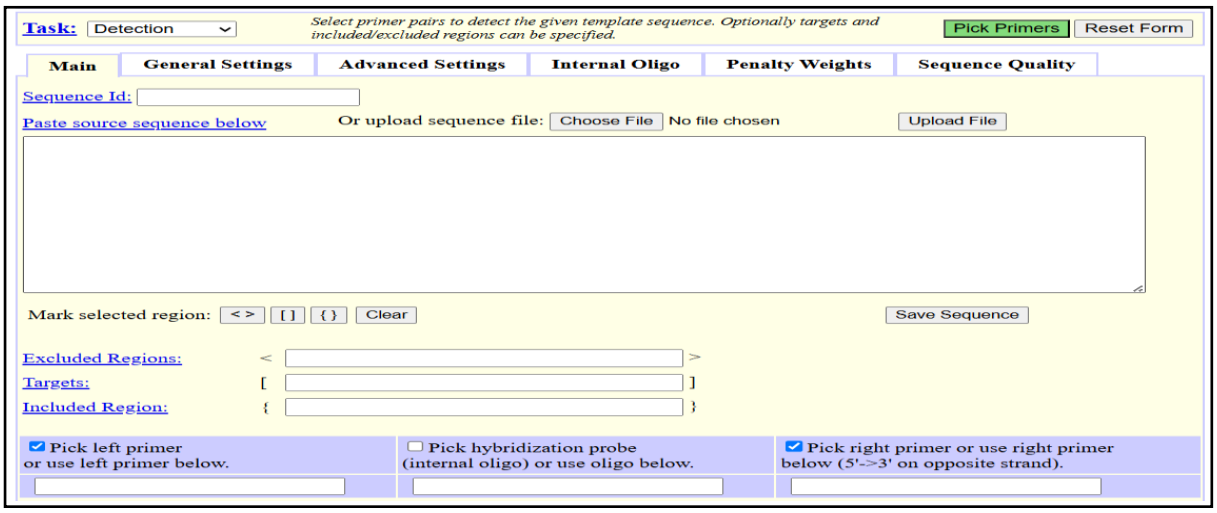


Fig.12: Primer3 tool layout used for designing gene-specific primers targeting conserved regions of COMT gene sequences in *B. tulda*

4.7.3. PCR amplification:

In order to confirm amplification efficiency of designed primers, PCR amplification was conducted to obtain bands having the desired length. Optimization of the annealing temperature was performed to ensure specificity. Each reaction was set up in a total volume of 50 µL, comprising genomic DNA (gDNA) template, forward and reverse primers, dNTPs, $MgCl_2$, 5× Taq buffer, and Taq DNA polymerase (Thermo Fisher Scientific). The reaction volume was adjusted to 50 µL using nuclease-free water (Table 1).

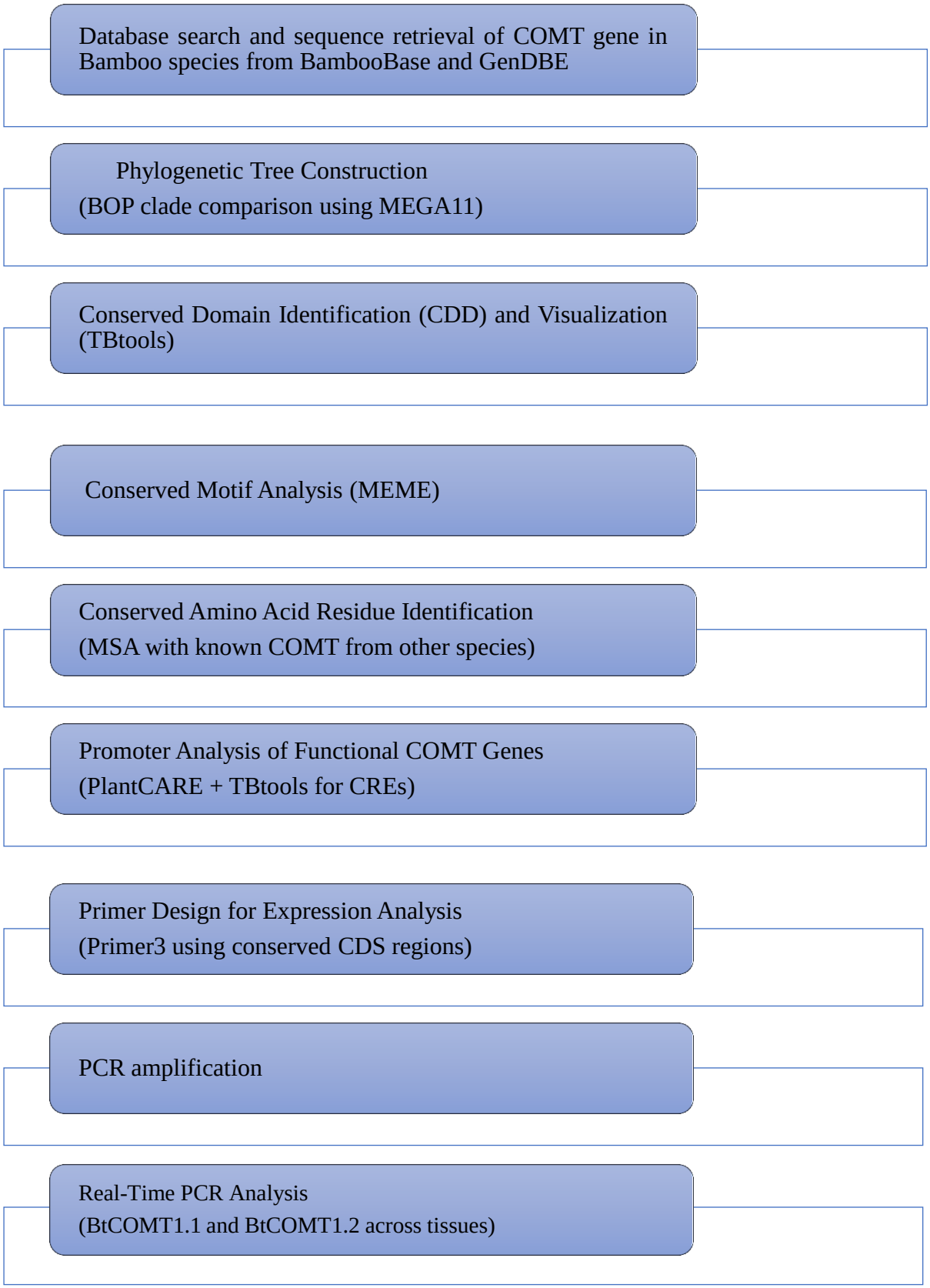
Table 1: PCR reaction components and their final concentrations used for amplification in a 50 µL reaction volume

PCR components	Stock	Working Concentration	Volume
Buffer	10X	1X	5
dNTPs	10mM	0.2Mm	1
Taq polymerase	5unit/ µL	1.25unit/50 µL	0.5
DNA template	180 µg/ µL	2 µg/ µL	0.6
Mgcl2	25mM	1mM	2
R primer	10µM	0.25 µM	1.25
F primer	10µM	0.25 µM	1.25

The amplification was performed in a thermal cycler using the following conditions: an initial denaturation at 95°C for 3 minutes, followed by 29 cycles of denaturation at 95°C for 30 seconds, annealing temperature of the primers are in gradients (depending on the melting temperature of both BtCOMT1.1 and BtCOMT1.2 primers), and extension at 72°C for 45 seconds. A final extension was carried out at 72°C for 5 minutes to ensure complete synthesis of the PCR products.

- The amplified products were separated on a 2% (w/v) agarose gel containing ethidium bromide and visualized under UV light. Amplification success was confirmed by the presence of a specific band at the expected size (~2kb).

Workflow of Methodology:


Database search and sequence retrieval of COMT gene in Bamboo species from BambooBase and GenDBE
Phylogenetic Tree Construction
(BOP clade comparison using MEGA11)
Conserved Domain Identification (CDD) and Visualization (TBtools)
Conserved Motif Analysis (MEME)
Conserved Amino Acid Residue Identification
(MSA with known COMT from other species)
Promoter Analysis of Functional COMT Genes
(PlantCARE + TBtools for CREs)
Primer Design for Expression Analysis
(Primer3 using conserved CDS regions)
PCR amplification
Real-Time PCR Analysis
(BtCOMT1.1 and BtCOMT1.2 across tissues)

5. RESULTS AND DISCUSSION

5.1. Retrieval of all relevant sequences retrieved from sequenced Bamboo genomes

Total of 34 COMT gene homologs from 12 species (Table 2) were obtained from the all-bamboo genome. The amino acid length of COMTs ranged from 260 to 360. All 34 sequences obtained from BLAST searches on Bamboobase and GenDBE have e-values = 0 and 100% query coverage.

Table 2: List of bamboo species with corresponding *COMT* gene family members and their accession numbers retrieved from databases

Species	Accession Number	COMT family	
Ampelocalamus luodianensis	Alu08Dg12340.2, Alu09Cg16220.2	2	BambooBase
Bonia amplexicaulis	Bam09Cg07910.1, Bam08Bg08560.1, Bam08Ag10130.1	3	
Raddia guianensis	Rgu08g15750.1	1	
Hsuehochloa calcarean	Hca08Dg02530.2, Hca09Cg07000.1	1	
Dendrocalamus sinicus	Dsi08Ag02350.1, Dsi09Cg09200.1, Dsi08Ag02340.1, Dsi08Bg01980.1	4	
Guadua angustifolia	Gan03Cg01560.2	1	
Melocanna baccifera	Mhu08Ag10640.1, Mhu09Cg08570.2	2	
Olyra latifolia	Ol08g20450.1	1	
Otatea glauca	Ogl09Cg05030.1, Ogl08Bg02510.1	2	
Phyllostachys edulis	Ped08Dg03190.1, Ped09Cg18050.1	2	
Rhipidocladum racemiflorum	Rhi09Cg13260.1, Rhi08Bg02380.1, Rhi07Bg17030.1	3	
Bambusa tulda	Btu09C.g01022t41,Btu08A.g01257t29, Btu08B.g00210t53,Btu08B.g00210t52, Btu08B.g00210t51,Btu08B.g00210t46, Btu08B.g00210t49,Btu08B.g00210t50, Btu08B.g00210t45,Btu08B.g00210t44, Btu08B.g00210t48, Btu08B.g00210t47	12	GenDBE

Out of the 12 COMT sequences identified in *Bambusa tulda*, two sequences (**Btu09C.g01022t41 and Btu08A.g01257t29**) are located on two different chromosomes 09C and 08A respectively (Fig.13 and 14).

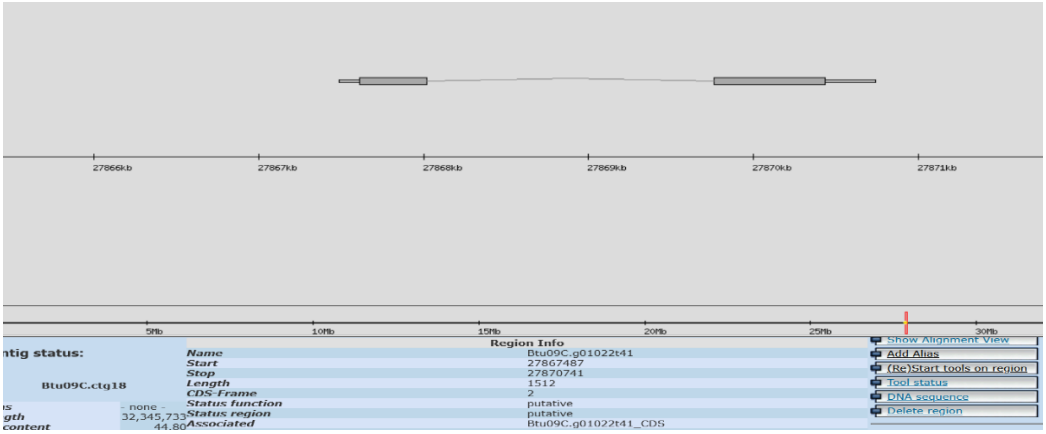


Fig.13: Identification of Btu090.g01022t41(BtCOMT1.1) using GenDBE

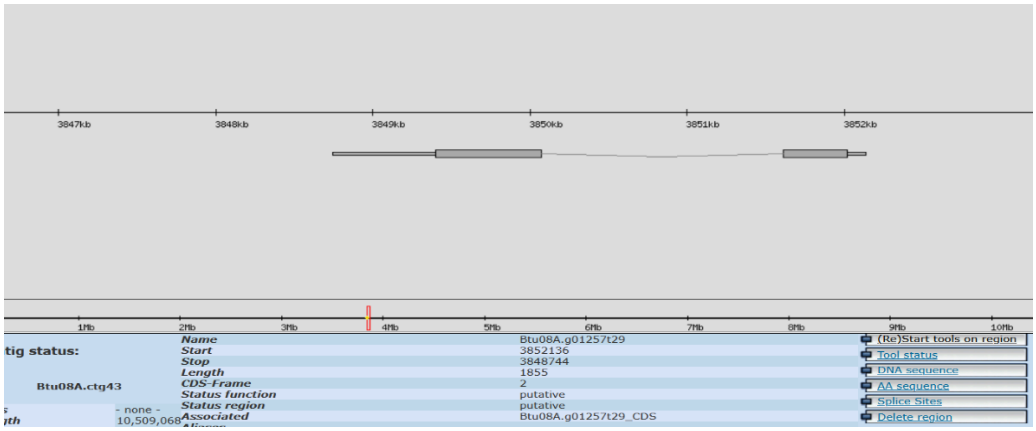


Fig.14: Identification of Btu08A.g01257t29(BtCOMT1.2) using GenDBE

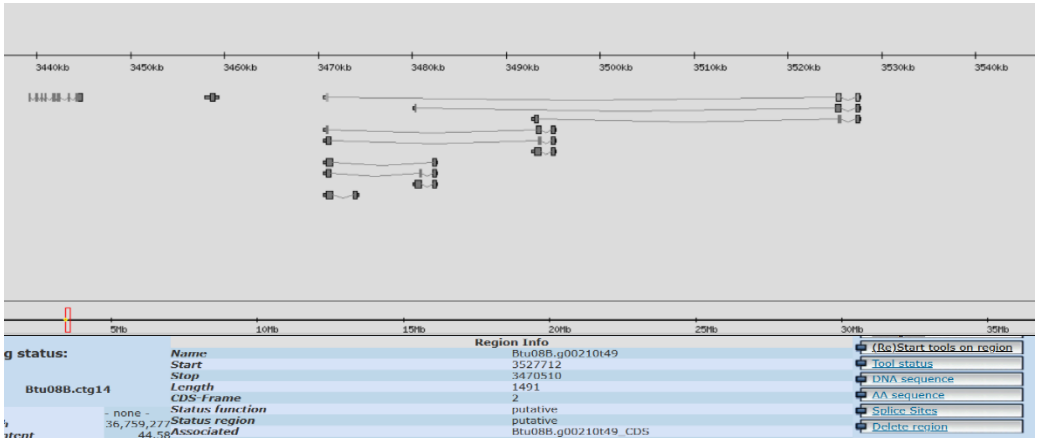


Fig.15: Identification of rest COMT sequences of *Bambusa tulda* using GenDBE

However, the remaining 10 COMT sequences (**Btu08B.g00210t53, Btu08B.g00210t52, Btu08B.g00210t51, Btu08B.g00210t46, Btu08B.g00210t49, Btu08B.g00210t50, Btu08B.g00210t45, Btu08B.g00210t44, Btu08B.g00210t48, Btu08B.g00210t47**) were found on a single region of a particular chromosome.

- From here, two sequences **Btu08B.g00210t46 (BtCOMT1.5) and Btu08B.g00210t45 (BtCOMT1.9)** were selected for further analysis based on the highest length of the CDS.

5.2. Phylogenetic analysis of all the obtained *COMT* genes in Bamboos:

A phylogenetic tree was initially constructed using 26 amino acid sequences (Fig.16) of the COMT gene from *Bambusa tulda* and 11 other bamboo species, including *Ampelocalamus luodianensis*, *Bonia amplexicaulis*, *Hsuehochloa calcarean*, *Dendrocalamus sinicus*, *Guadua angustifolia*, *Melocanna baccifera*, *Olyra latifolia*, *Otatea glauca*, *Phyllostachys edulis*, *Raddia guianensis*, and *Rhipidocladum racemiflorum*. The tree was generated using the Maximum Likelihood (ML) method in MEGA11, following multiple sequence alignment.

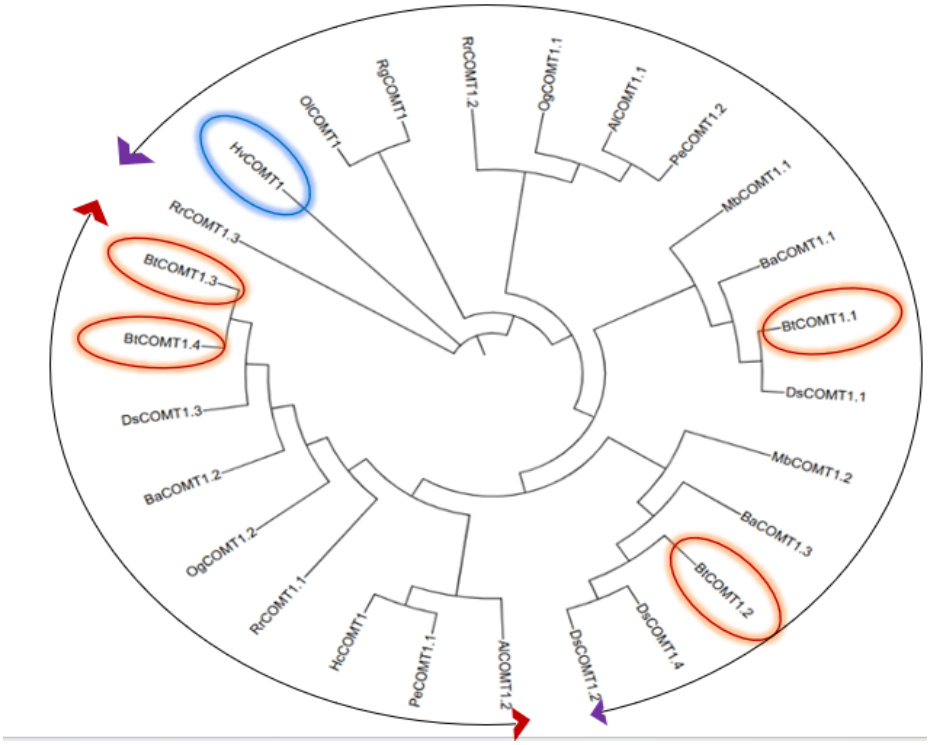


Fig.16: Maximum Likelihood (ML) phylogenetic tree constructed using 26 COMT amino acid sequences. The tree shows distribution of *B. tulda COMT* genes across multiple clades

- To obtain better resolution of the evolutionary relationships among *B. tulda* COMT genes, a second phylogenetic tree was constructed using COMT amino acid sequences from species within the BOP clade (Bambusoideae, Oryzoideae, and Pooideae). This group was selected because it includes species that are more closely related to bamboo. In the BOP-focused tree, the *B. tulda* COMT sequences clustered more closely with those from other Bambusoideae members (Fig.17), suggesting a shared evolutionary origin and potentially similar functions.

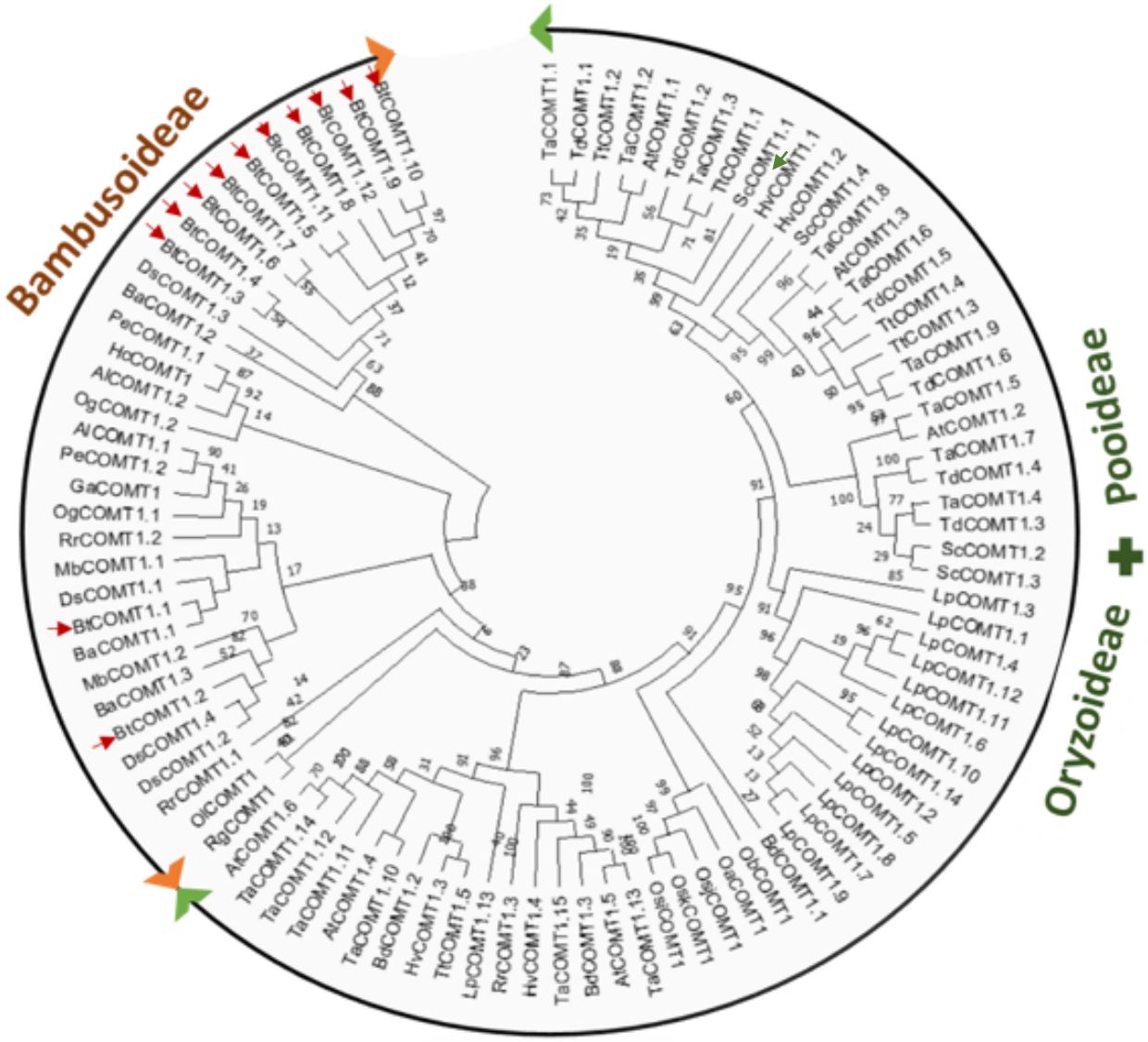


Fig.17: Maximum Likelihood (ML) phylogenetic tree based on COMT amino acid sequences from species in the BOP clade (Bambusoideae, Oryzoideae, and Pooideae). *Bambusa tulda COMT* genes cluster more closely with other Bambusoideae species

Phylogenetic comparison of COMT amino acid sequences revealed that *Bambusa tulda* shares a closer evolutionary relationship with other BOP clade grasses, consistent with previous findings in monocots (Shen *et al.,* 2013).

5.3. Identification of conserved signature domain in *COMT* protein sequences in Bamboos:

Conserved domain analysis was performed using 26 amino acid sequences of the deduced *COMT* protein from *Bambusa tulda* and 11 other bamboo species. The results revealed that all COMT proteins contain the highly conserved **C-terminal Methyltransferase domain (Methyltransf_2, PF00891)**, which includes the **SAM/SAH-binding pocket** and the **substrate-binding site** essential for catalytic activity by transferring a methyl group from SAM to lignin precursors. In addition, many of the sequences also exhibited the **N-terminal Dimerization domain (PF08100)** (Table 3)**,** which plays an important role in conferring structural stability and protein-protein interaction. Notably, all analysed COMT proteins were found to contain both conserved domains (Fig.18), confirming their likely functionality as S-adenosyl-L-methionine-dependent O-methyltransferases.

Table 3: Conserved domain prediction table showing the presence of domains in COMT proteins across selected bamboo species

Query	Hit type	PSSM-ID	From	To	E-Value	Bitscore	Accession	Short name	Incomplete	Superfamily
Q#1 - >RgCOMT1	specific	395719	137	342	2.71707e-99	292.387	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	7.33458e-17	73.3813	pfam08100	dimerization	-	cl06920
Q#2 - >HcCOMT1	specific	395719	137	342	1.04584e-102	301.247	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	5.40234e-18	76.8481	pfam08100	dimerization	-	cl06920
Q#3 - >PeCOMT1.1	specific	395719	137	342	1.02327e-102	301.247	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	1.85098e-18	78.0037	pfam08100	dimerization	-	cl06920
Q#4 - >PeCOMT1.2	specific	395719	137	341	3.6304e-103	302.402	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	2.04798e-18	78.0037	pfam08100	dimerization	-	cl06920
Q#5 - >GaCOMT1	specific	395719	65	269	1.4528e-103	300.476	pfam00891	Methyltransf_2	-	cl17173
Q#6 - >OlCOMT1	specific	395719	137	342	9.06456e-102	298.935	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	1.04084e-17	76.0777	pfam08100	dimerization	-	cl06920
Q#7 - >AlCOMT1.1	specific	395719	137	341	5.93106e-103	302.017	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	3.91062e-18	77.2333	pfam08100	dimerization	-	cl06920
Q#8 - >AlCOMT1.2	specific	395719	136	341	4.81129e-103	302.017	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	30	81	1.64167e-18	78.0037	pfam08100	dimerization	-	cl06920
Q#9 - >OgCOMT1.1	specific	395719	137	342	4.61892e-104	304.713	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	3.53487e-18	77.2333	pfam08100	dimerization	-	cl06920
Q#10 - >OgCOMT1.2	specific	395719	137	342	9.78368e-102	298.935	pfam00891	Methyltransf_2	-	cl17173
	specific	400439	31	82	2.31219e-18	77.6185	pfam08100	dimerization	-	cl06920

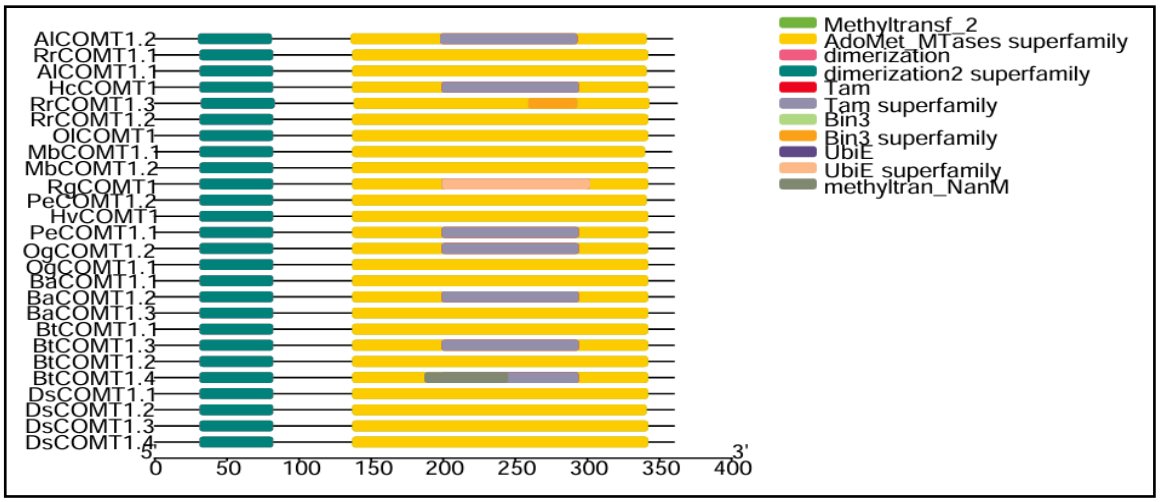


Fig.18: Graphical representation of conserved domains in COMT proteins, highlighting the structural arrangement of associated domains

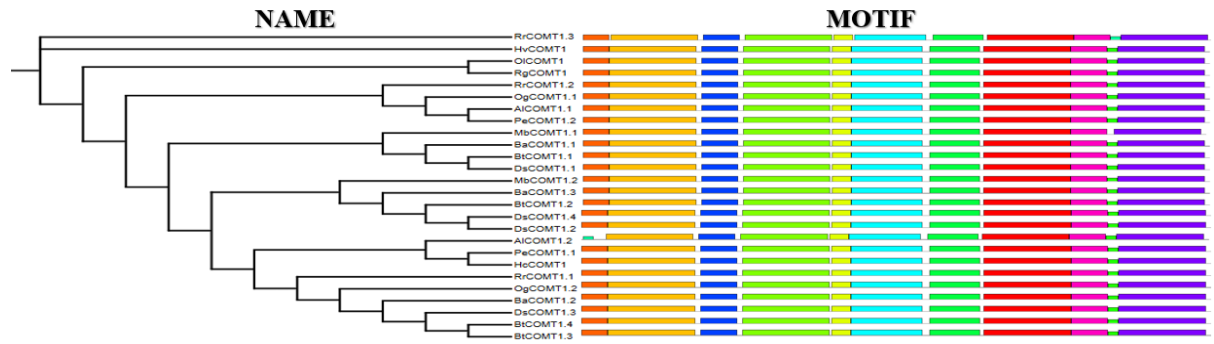
The identification of both Methyltransf_2 and Dimerization domains in all COMT proteins from *Bambusa tulda* strongly supports their role as catalytically active enzymes involved in S-lignin biosynthesis (Zhao *et al*., 2013). The Methyltransf_2 domain provides the methyltransferase activity necessary for transferring methyl groups from SAM to lignin precursors, while the Dimerization domain ensures proper protein folding and structural stability through homodimer formation (Joshi & Chiang *et al*. 1998). Their conserved presence not only validates the functional identity of the predicted COMT proteins but also highlights their potential as reliable molecular targets for lignin modification strategies aimed at reducing biomass recalcitrance in bamboo.

5.4. Conserved motif analysis of *COMT* genes in bamboos:

Motif analysis was conducted using 26 amino acid sequences of the *COMT* gene from *Bambusa tulda* and 11 other bamboo species to identify conserved regions that may contribute to proteins' structural or functional roles. The number of motifs to be identified was set to 12 (Fig.19), with motif lengths ranging from 6 to 50 amino acids. The results showed that most sequences contained 10 to 12 conserved motifs, with several motifs being highly consistent across all species analysed.

Fig.19: Conserved motif logos of COMT amino acid sequences from *B. tulda* and other bamboo species, identified using the MEME Suite

- The identified motifs were uniformly distributed in nearly all COMT proteins (Fig.20), suggesting a high level of conservation and potential functional relevance. In *Bambusa tulda*, all COMT proteins exhibited conserved motif arrangements similar to those in species like *Phyllostachys edulis* and *Dendrocalamus sinicus*, indicating that the *COMT* gene family in *Bambusa tulda* shares evolutionary and functional similarity with other members of the Bambusoideae clade.



Motif	Symbol	Motif Consensus
1.		HVISEAPPFPGVQHVGGNMFEKVPSGDAILMKWILHDWSDEHCATLLKNC
2.		YGMTAFEYHGTDPRFNRVFNEGMKNHSIIITKKLLEFYTGF
3.		EGKDGRLARRYGPAPVCKWLTPNEDGVSMAALALMNQDKVLMESWYYLKD
4.		QGVFHVDMIMLAHNPGGKERYEREFEELARGAGFAGVKATYIYANAWAIE
5.		ACMYAMQLASSSILPMTLKNAIELGLLEILVGAGGKALSPAEVAARLPST
6.		TLVDVGGGIGATLHAITSKYPHIKGINFD
7.		DAPAMVDRMLRLLASYNVVSC
8.		YDALPAHGKVIIVECILPVNP
9.		MGSTAADMAAAADEE
10.		VLDGGIPFNKA
11.		EATPKA
12.		GVTNRD

Fig.20: Conserved motif distribution among all COMT proteins, where each coloured box represents a specific conserved motif identified by MEME

- The consistent motif patterns likely represent functionally important regions such as methyl group donor binding sites (SAM-binding motifs) and catalytic sites essential for methyltransferase activity. These findings support the reliability of the identified COMT sequences and their potential role in S-lignin biosynthesis. The structural conservation observed through motif analysis complements the conserved domain findings and reinforces the predicted functional activity of COMT proteins in *Bambusa tulda* and other bamboo species (Zhong & Ye, 2015; Wang *et al*., 2022).

5.5. Analysis on conserved amino acid residues of *COMT* genes:

Multiple sequence alignment (MSA) of 26 COMT protein sequences from *Bambusa tulda* and 11 other bamboo species revealed highly conserved residues, particularly within the central region of the proteins. These conserved residues were consistently aligned across all sequences, especially within regions corresponding to the known catalytic and binding domains of S-adenosyl-L-methionine (SAM)-dependent O-methyltransferases (Fig.21). Residues such as glycine (G), valine (V), aspartate (D), and histidine (H) were found to be conserved at key positions, suggesting a functional role in enzymatic activity. These conserved stretches are characteristic of SAM-binding loops and catalytic cores commonly associated with methyl group transfer in lignin biosynthesis.

The conservation of these residues is functionally significant and reflects evolutionary pressure to maintain methyltransferase activity across bamboo species. Previous structural studies have shown that conserved glycine-rich regions form part of the SAM-binding pocket, while aromatic and polar residues interact with lignin precursors such as 5-hydroxyferulic acid (Zubieta *et al*., 2001; Joshi & Chiang, 1998). These residues form the core of the Methyltransf_2 domain (PF00891), which is critical for catalysis. Their presence in all aligned sequences indicates that COMT proteins in *B. tulda* and other bamboos retain their functional identity as O-methyltransferases.

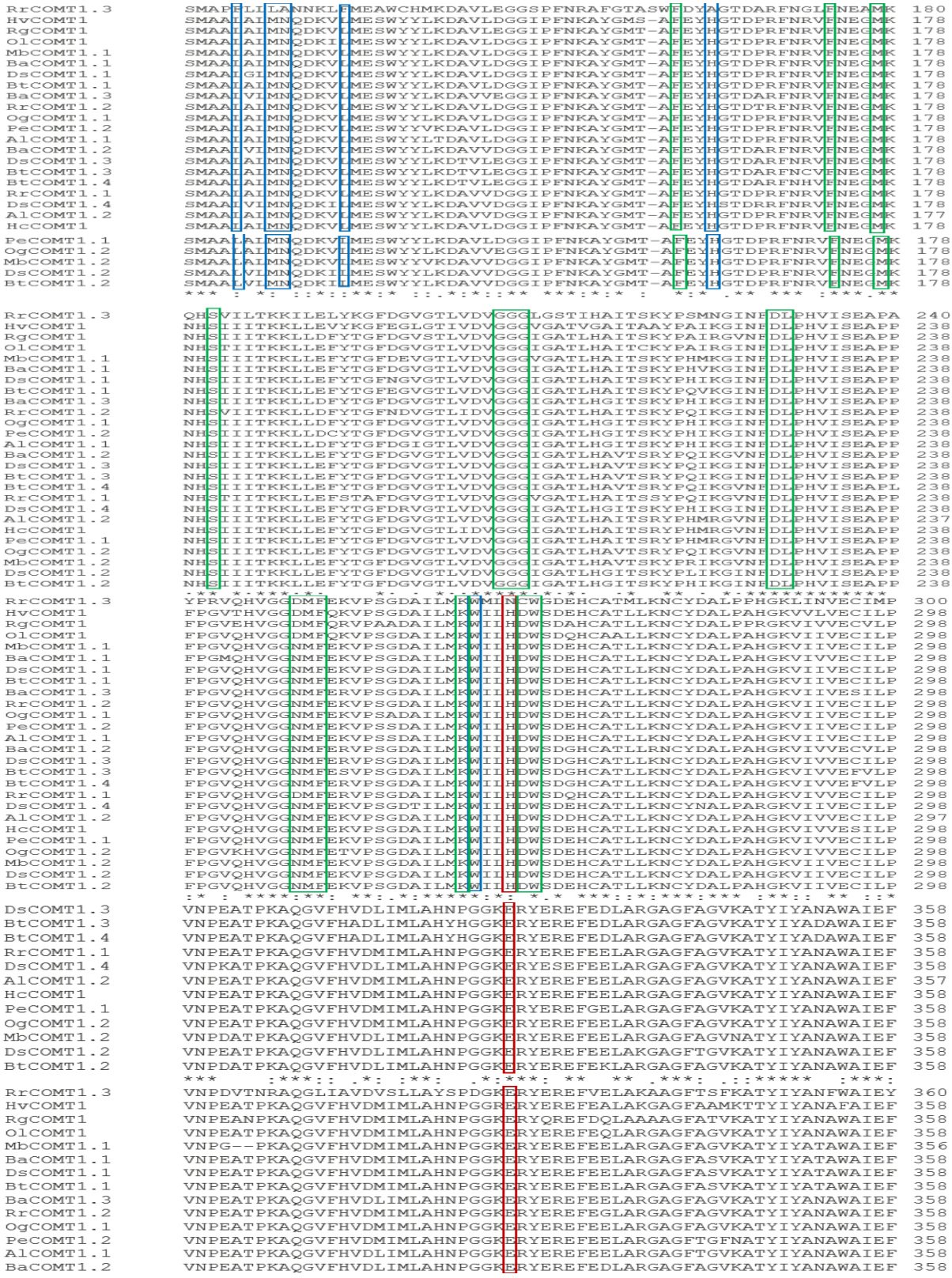


Fig. 21: Multiple Sequence Alignment of Bamboos with known *COMT* gene of *Hordeum vulgare* were performed and conserved amino acid residues were identified.

Blue - Substrate binding Green - SAM binding Orange - Catalytic residues

Comparative studies on species like *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays* have similarly reported that conserved residues within COMTs are essential for maintaining enzymatic function and substrate specificity (Halpin *et al.*, 1998; Shen *et al.*, 2013). The results of this study align with such findings and validate the predicted bamboo COMT sequences as functional candidates involved in syringyl lignin biosynthesis. Additionally, conservation of this residue supports their potential as molecular targets for future genetic modification to improve lignin composition in bamboo, particularly for bioenergy applications (Biswas *et al.*, 2022).

Collectively, these analyses provide a solid molecular basis for selecting key COMT genes that can be further explored through promoter and expression profiling to support lignin modification and biomass improvement in bamboo.

- Among the four COMT sequences of *Bambusa tulda*, which were selected for this analysis, two sequences are located on separate chromosomes, each possessing distinct upstream promoter regions. In contrast, the remaining two sequences are situated within the same chromosomal region. Therefore, promoter analysis was carried out only for the two COMT sequences (*BtCOMT1.1* and *BtCOMT1.2*) that possess independent promoter regions.

5.6. Analysis on promoter and conserved Cis-regulatory elements of *COMT* genes in *Bambusa tulda*:

Cis-regulatory elements (CREs) in plants are known to bind transcription factors that regulate their expression at various developmental stages and environmental conditions. To explore the regulatory mechanisms of the *COMT* gene in *Bambusa tulda*, 2000 base pairs upstream promoter regions of two functional *COMT* genes - *BtCOMT1.1 and BtCOMT1.2* were retrieved from GenDBE and further analysed using the PlantCARE database (Fig.22).

- The identified CREs were subsequently classified according to their functional categories, as presented in the accompanying tables (Table 1 and Table 2).

Table 4: A series of CREs found in the promoter regions of the *BtCOMT1.1*

Cis regulatory element	Sequence	Function	Copy
CAAT-box	CAAAT	Developmental processes	3
	CAAT		11
	CCAAT		1
GT1-motif	GGTTAA	Light responsive	2
	GGTTAAT		1
Box 4	ATTAAT		1
I-box	GATAA		3
MYB	CAACCA	Stress responses	1
LTR	CCGAAA		1
STRE	AGGGG		1
ARE	AAACCA		1
	TGGTTT		2
MYC	CATGTG		1
	CAATTG		1
W-box	TTGACC		1

The promoter analysis of *BtCOMT1.1* revealed multiple cis-regulatory elements associated with developmental, environmental, and stress-responsive regulation. The presence of CAAT-box variants indicates involvement in core transcriptional activity. Several light-responsive elements suggest that *BtCOMT1.1* may be regulated by light conditions, aligning with known light-dependent regulation of lignin biosynthesis genes. Additionally, stress-related elements indicate responsiveness to abiotic stress signals, while the presence of LTR and ARE elements suggests potential regulation under cold and anaerobic conditions. These findings imply that *BtCOMT1.1* expression may be modulated by both developmental cues and environmental stresses, contributing to its role in lignin biosynthesis and plant adaptation.

Table 5: A series of CREs found in the promoter regions of the *BtCOMT1.2*

Cis regulatory element	Sequence	Function	Copy no.
G - box	CAATTG	Light response	2
Sp1	GGGCGG		1
Box 4	ATTAAT		1
I-box	GATAA		3
CGTCA - motif	CGTCA	Stress response	1
MYB	CAACCA		2
MYC	CATTTG		2
	CAATTG		2
MBS	CAACTG		1
CAAT-box	CAAAT	Developmental response	3
	CAAT		11
	CCAAT		3
RY-element	CATGCATG		1
	CATGCA		2
MYB recognition site	CCGTTG	Lignin-specific gene activation	2
AC-II	ACCAACC	MYB binding	2

The promoter region of *BtCOMT1.2* contains a wide range of cis-regulatory elements associated with light responsiveness, stress signalling, developmental processes, and lignin-specific regulation. Elements such as the G-box, I-box, and Box 4 suggest that light may play a regulatory role in *BtCOMT1.2* expression. The presence of multiple stress-related motifs indicates potential involvement in environmental stress responses. Notably, the identification of the **MYB recognition site** and **AC-II element**, both linked to lignin-specific gene activation, highlights the gene's likely functional role in lignin biosynthesis (Jianli *et al.,* 2009). The presence of RY-elements and CAAT-box variants further supports regulation during seed development and general transcriptional control. Altogether, these CREs suggest that *BtCOMT1.2* is transcriptionally responsive to both environmental and developmental signals, reinforcing its role in adaptive lignification in *Bambusa tulda*.

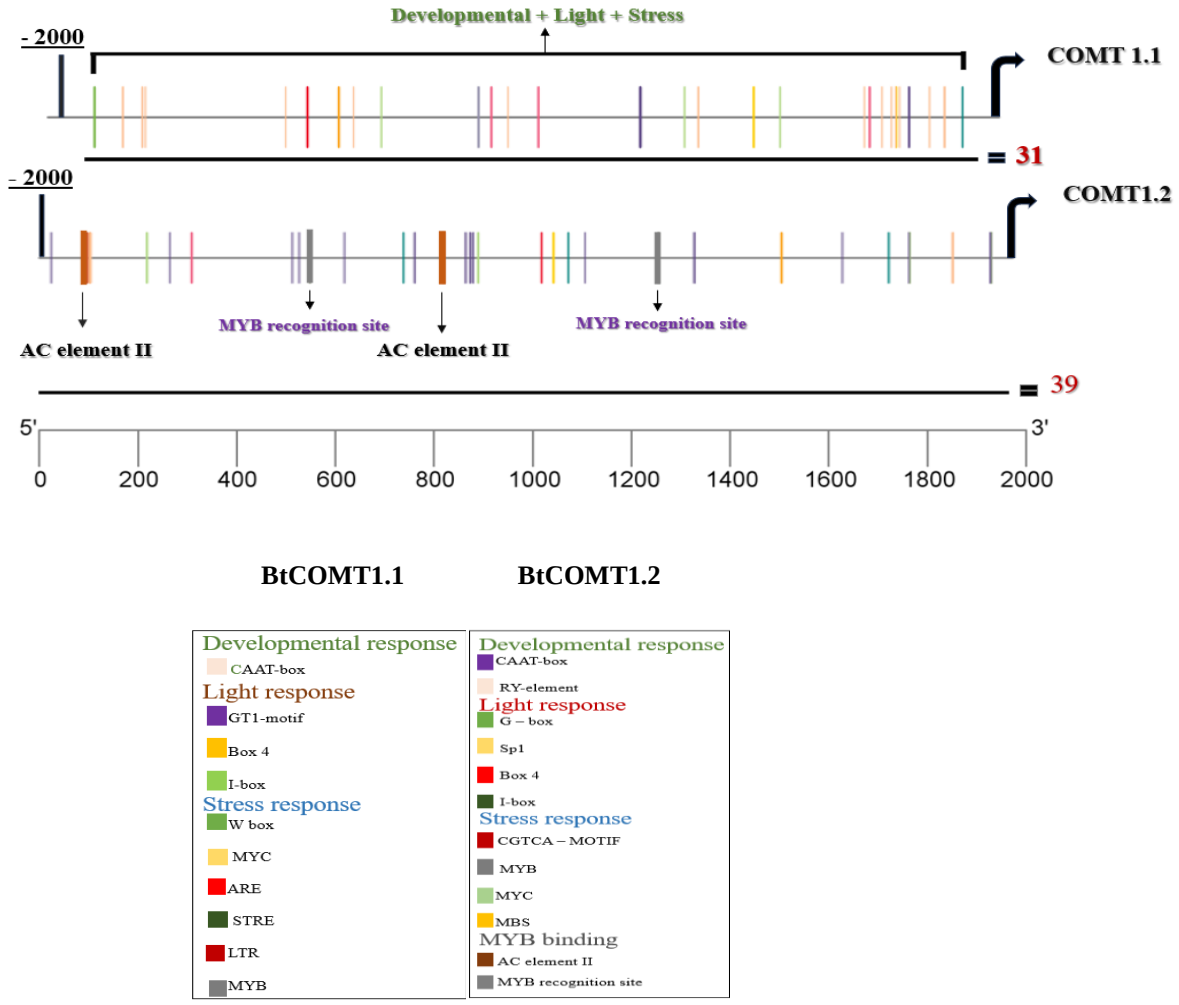


Fig.22: Cis-regulatory elements in the 2kb upstream regions of *COMT* gene in B. tulda. Different colour represents different cis elements

5.7. Expression Analysis of *COMT* Genes in the Stem of B. *tulda*:

5.7.1. Primer designing:

A pair of gene-specific primers was designed (Table 6) from each of the two *COMT* gene sequences of *Bambusa tulda* (*BtCOMT1.1* and *BtCOMT1.2*) to evaluate their expression by real-time PCR. The primers were selected from conserved regions within the respective CDS.

Table 6: Gene-specific primers designed from *COMT* genes in ***Bambusa tulda***

Gene Id	Primer direction	Primer seq. (5′–3′)	Length (bp)	Tm (°C)	GC content (%)
BtCOMT1.1	Forward (F)	CTCACCCCCAACGAGGAT	18	60.9	61
	Reverse (R)	GATGCCGCCGTCAAGG	16	62.5	68.8
BtCOMT1.2	Forward (F)	CCTCATGAACCAGGACAAGA	20	58.6	50
	Reverse (R)	TTCATGCCCTCGTTGAAC	18	58.1	50

5.7.2. PCR amplification:

Gradient PCR amplification was performed using gene-specific primers on a BIORAD S1000 thermal cycler. The reaction conditions were optimized to determine the annealing temperature that ensures specific amplification of the target gene regions.

- PCR analysis using 2% agarose gel electrophoresis followed by ethidium bromide staining revealed that the intensity and size of the amplified bands depended on the specificity of the gene-specific primers used.
- The amplification product of BtCOMT1.1 using genomic DNA as template showed a prominent band at ~2 kb at an annealing temperature of 65 °C, while BtCOMT1.2 produced clear bands at annealing temperatures of 51 °C and 52 °C, indicating successful amplification of the functional *COMT* gene copies.

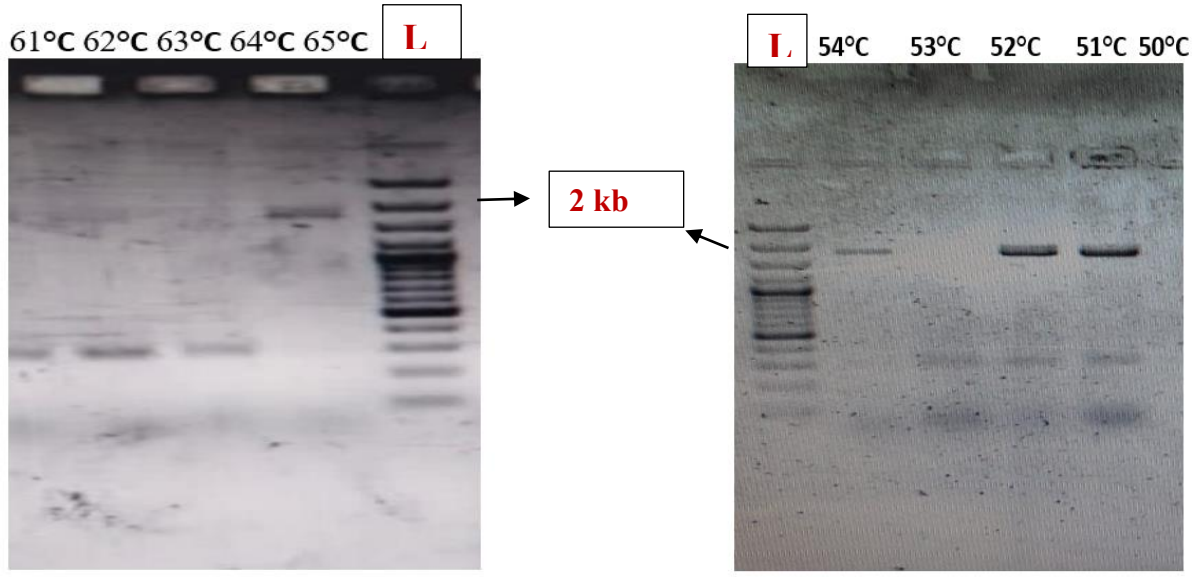


Fig.23: Gradient PCR showing optimal amplification of *BtCOMT1.1* at 65 °C (left) and *BtCOMT1.2* at 51–52 °C (right) using gene-specific primers

CONCLUSION

The comprehensive identification and characterization of the *COMT* gene family in *Bambusa tulda* presented in this study contribute significantly to understanding the molecular regulation of lignin biosynthesis in bamboo. Given *Bambusa tulda's* high cellulose content, rapid growth, and commercial relevance, improving its biomass properties for biofuel production and other industrial uses holds considerable promise. Among the enzymes involved in the phenylpropanoid pathway, COMT plays a critical role in the methylation of lignin precursors, particularly in the synthesis of syringyl lignin units, which affect lignin structure, solubility, and biomass digestibility.

This study applied a genome-wide approach to identify 12 COMT gene sequences in *B. tulda*. Phylogenetic analysis using COMT sequences from related BOP clade species revealed evolutionary patterns consistent with functional divergence. Conserved domain and motif analyses confirmed the presence of key functional regions such as the Methyltransf_2 and Dimerization domains in most *BtCOMT* proteins, validating their roles in catalyzing methylation steps during lignin formation. Furthermore, multiple sequence alignment helped in identifying conserved amino acid residues, strengthening the annotation of true functional COMT protein.

Promoter analysis of two *COMT* genes, BtCOMT1.1 and BtCOMT1.2, revealed cis-regulatory elements responsive to light, hormones, and various stresses, suggesting dynamic regulation of these genes in response to developmental and environmental signals. Expression analysis using real-time PCR across differ ent developmental stages of *B. tulda* stem tissue demonstrated stage-specific expression patterns, implying their involvement in lignin deposition during stem maturation.
Altogether, these findings provide a strong foundation for further functional validation and genetic engineering efforts aimed at modifying lignin biosynthesis in bamboo. By manipulating *COMT* gene expression or function, it may be possible to reduce biomass recalcitrance and improve saccharification efficiency—thereby enhancing the suitability of *Bambusa tulda* as a sustainable feedstock for second-generation bioethanol production. Future studies should focus on experimental validation using gene editing tools and field-level evaluation to translate these molecular insights into applied outcomes for bioenergy and industrial applications.

FUTURE PROSPECTS

The identification and characterization of *COMT* gene family members in *Bambusa tulda* provide a foundational understanding of their potential roles in lignin biosynthesis and regulation. Building upon these findings, the following future research directions are proposed:

- **Expression Profiling at Developmental Stages:**

  A detailed expression analysis of target COMT genes should be conducted across various developmental stages of *Bambusa tulda*, particularly focusing on stem tissues, to understand their temporal regulation and contribution to lignification.

- **Tissue specific expression analysis:**

  Investigating the tissue-specific expression of COMT genes in roots, leaves, nodes, internodes, and vascular tissues will provide insights into their spatial regulation and potential functional diversification in different plant organs.

- **Experimental Validation of Cis-Regulatory Elements:**

  The predicted cis-regulatory elements identified in the promoter regions of COMT genes may be experimentally validated by generating promoter deletion constructs and transforming them into model systems. Comparative expression analysis of promoter lines with and without specific cis-elements will help confirm their functional roles in gene regulation.

- **Functional Characterization via Transgenic Overexpression:**

  Selected *COMT* genes from *Bambusa tulda* can be functionally characterized by overexpressing them in *Arabidopsis thaliana*. These transgenic lines can be analyzed for alterations in lignin composition, biomass digestibility, and S/G ratio, thereby validating the functional role of bamboo-derived COMTs in lignin biosynthesis.

8. REFERENCES

1. Liang, S., Xu, S., Qu, D., Yang, L., Wang, J., Liu, H., Xin, W., Zou, D., & Zheng, H. (2022). *Identification and functional analysis of the caffeic acid O-methyltransferase (COMT) gene family in rice (Oryza sativa L.)*. *International Journal of Molecular Sciences, 23*(15), 8491. https://doi.org/10.3390/ijms23158491
2. Lei, D., Chen, Y., Li, Y., Hu, Y., Zhang, J., & Wang, L. (2024). *Genome-wide identification of COMT gene family in maize and its function in response to light*. *Biochemical Genetics.* https://doi.org/10.1007/s10528-024-10942-y
3. Fornalé, S., Sonbol, F. M., Maes, T., Capellades, M., Puigdomènech, P., Rigau, J., & Caparrós-Ruiz, D. (2006). Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors. *Plant Molecular Biology*, 62(6), 809–823. https://doi.org/10.1007/s11103-006-9058-2
4. Daly, P., McClellan, C., Maluk, M., Oakey, H., Lapierre, C., Waugh, R., Stephens, J., Marshall, D., Barakate, A., Tsuji, Y., Goeminne, G., Vanholme, R., Boerjan, W., Ralph, J., & Halpin, C. (2019). RNAi-suppression of barley caffeic acid O-methyltransferase modifies lignin despite redundancy in the gene family. *Plant Biotechnology Journal*, 17(3), 594–607. https://doi.org/10.1111/pbi.13001
5. Van Acker, R., Déjardin, A., Desmet, S., Hoengenaert, L., Vanholme, R., Morreel, K., Laurans, F., Kim, H., Santoro, N., Foster, C. E., Goeminne, G., Grabber, J. H., Ralph, J., Pilate, G., & Boerjan, W. (2017). Different routes for conifer- and sinapyl alcohol-derived lignin biosynthesis in poplar and Arabidopsis. *Biotechnology for Biofuels*, 10, 144. https://doi.org/10.1186/s13068-021-02026-5
6. Boerjan, W., Ralph, J., & Baucher, M. (2003). Lignin biosynthesis. *Annual Review of Plant Biology*, 54, 519–546. https://doi.org/10.1146/annurev.arplant.54.031902.134938
7. Rahaman, A., Rangan, L., & Das, A. K. (2023). Comparative evaluation of lignocellulosic biomass from bamboo species for sustainable bioethanol production. *Biomass Conversion and Biorefinery*, 13(2), 1457–1470. https://doi.org/10.1007/s13399-022-02255-7

8. Vanholme, R., Demedts, B., Morreel, K., Ralph, J., & Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiology*, 153(3), 895–905. https://doi.org/10.1104/pp.110.155119
9. Zhao, Q., & Dixon, R. A. (2011). Transcriptional networks for lignin biosynthesis: More complex than we thought? *Trends in Plant Science*, 16(11), 659–668. https://doi.org/10.1016/j.tplants.2011.08.003
10. Barakat, A., Choi, A., Yassin, N. B. M., Park, J. S., Sun, Z., & Carlson, J. E. (2011). Comparative genomics and evolutionary analyses of the O-methyltransferase gene family in *Populus*. *Gene*, 479(1–2), 37–46. https://doi.org/10.1016/j.gene.2011.01.010
11. Agarwal, P. K., Agarwal, P., Reddy, M. K., & Sopory, S. K. (2006). Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Reports*, 25(12), 1263–1274. https://doi.org/10.1007/s00299-006-0204-8
12. Biswas, A., Sahoo, S., Ghosh, A., & Mandal, S. (2022). Genetic and biochemical evaluation of *Bambusa tulda* for biofuel production. *Renewable Energy*, 190, 345–354. https://doi.org/10.1016/j.renene.2022.03.019
13. Goujon, T., Sibout, R., Eudes, A., MacKay, J. J., & Jouanin, L. (2003). Genes involved in lignin biosynthesis. *Plant Biology*, 5(2), 120–133. https://doi.org/10.1055/s-2003-39596
14. Halpin, C., Holt, K., Chojecki, J., Oliver, D., Chabbert, B., Monties, B., & Edwards, K. (1998). Brown-midrib maize (bm1–bm5): A model system for understanding the genetic and molecular basis of lignification. *The Plant Journal*, 14(5), 545–553. https://doi.org/10.1046/j.1365-313X.1998.00165.x
15. Shen, H., Poovaiah, C. R., Ziebell, A., Tschaplinski, T. J., Pattathil, S., Gjersing, E., ... & Chen, F. (2013). Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnology for Biofuels*, 6, 71. https://doi.org/10.1186/1754-6834-6-71
16. Vidgren, J., Svensson, L. A., & Liljas, A. (1994). Crystal structure of catechol O-methyltransferase. *Nature Structural Biology*, 1(7), 439–446. https://doi.org/10.1038/nsb0794-439
17. Lee, J. H., Won, H. J., Tran, P. H. N., Lee, S. M., Kim, H. Y., & Jung, J. H. (2021). Improving lignocellulosic biofuel production by CRISPR/Cas9-mediated lignin modification in barley. *GCB Bioenergy*, 13(4), 742–752. https://doi.org/10.1111/gcbb.12808

18. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
19. Liu, Y., Wang, Y., Pei, J., Li, Y., & Sun, H. (2021). Genome-wide identification and characterization of COMT gene family during the development of blueberry fruit. *BMC Plant Biology, 21*,1-16. https://doi.org/10.1186/s12870-020-02767-9
20. Lu, N., Ma, W., Han, D., Liu, Y., Wang, Z., Wang, N., Yang, G., Qu, G., Wang, Q., Zhao, K., & Wang, J. (2019). Genome-wide analysis of the *Catalpa bungei* caffeic acid O-methyltransferase (COMT) gene family: Identification and expression profiles in normal, tension, and opposite wood. *PeerJ, 7*, e6520. https://doi.org/10.7717/peerj.6520
21. Zhang, X., Chen, B., Wang, L., Ali, S., Guo, Y., Liu, J., Wang, J., Xie, L., & Zhang, Q. (2021). Genome-wide identification and characterization of caffeic acid O-methyltransferase gene family in soybean. *Plants, 10*(12), 2816. https://doi.org/10.3390/plants10122816
22. Van Bel, M., Silvestri, F., Weitz, E. M., Kreft, Ł., Botzki, A., Coppens, F., Vandepoele, K., & Van de Peer, Y. (2022). PLAZA 5.0: Extending the scope and power of comparative, evolutionary and functional genomics in plants. *Nucleic Acids Research*, 50(D1), D1468–D1474. https://doi.org/10.1093/nar/gkab965
23. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. https://doi.org/10.1093/molbev/msab120
24. Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., ... & Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research*, 48(D1), D265–D268. https://doi.org/10.1093/nar/gkz991
25. Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 13(8), 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009
26. Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49. https://doi.org/10.1093/nar/gkv416
27. Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, 27(1), 135–145. https://doi.org/10.1002/pro.3290

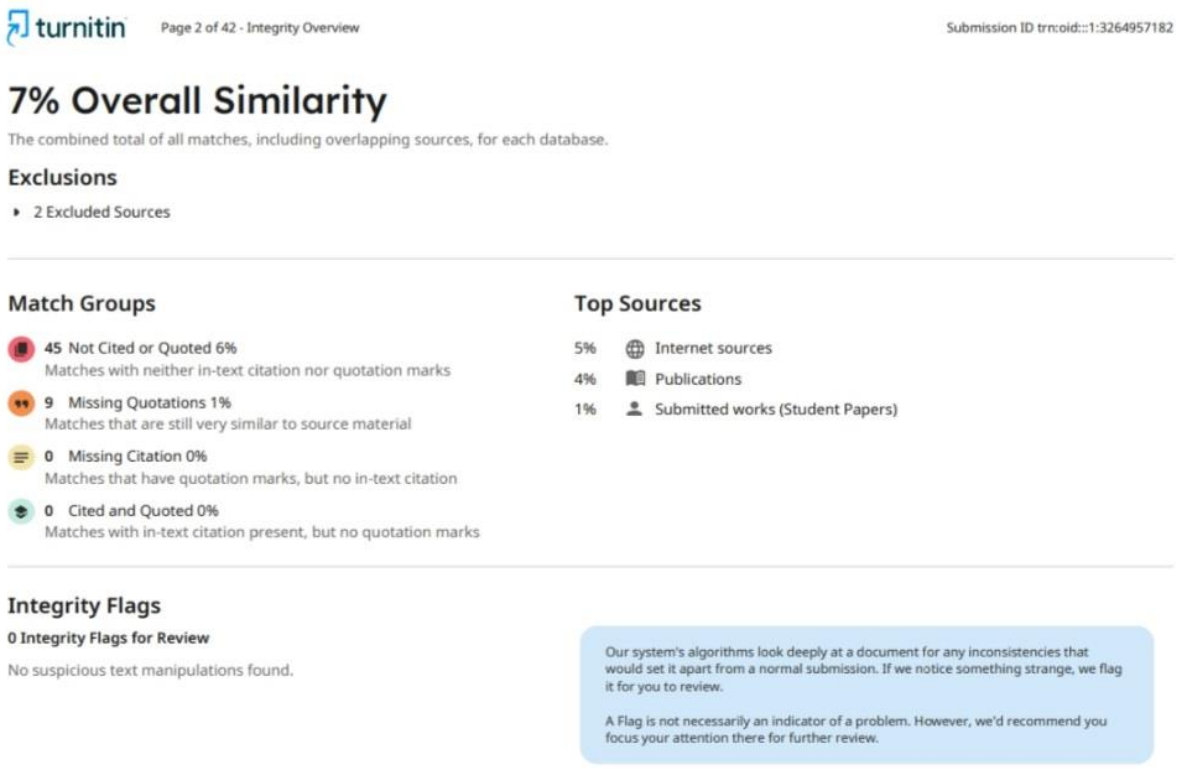
28. Zhong, R., Lee, C., Zhou, J., McCarthy, R. L., & Ye, Z. H. (2008). A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *The Plant Cell, 20*(10), 2763–2782. https://doi.org/10.1105/tpc.108.061325
29. Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., ... & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 30(1), 325–327. https://doi.org/10.1093/nar/30.1.325
30. Dutta, S., Deb, A., Biswas, P., Chakraborty, S., Guha, S., Mitra, D., Geist, B., Schäffner, A. R., & Das, M. (2021). Identification and functional characterization of two bamboo FD gene homologs having contrasting effects on shoot growth and flowering. *Scientific Reports, 11*, 7849. https://doi.org/10.1038/s41598-021-87491-6
31. Chakraborty, S., Dutta, S., Biswas, P., & Das, M. (2019). Identification of candidate reference genes in tropical bamboos stable across species, tissues, and developmental stages. *Biologia Plantarum, 63*(1), 253–261. https://doi.org/10.32615/bp.2019.029
32. Liu, C., Chen, Y., Zhu, Q., Lin, Z., Yu, Y., Zhang, T., & Wu, L. (2021). Identification and analysis of conserved motifs in the COMT gene family among angiosperms. *Plant Gene, 27*, 100310. https://doi.org/10.1016/j.plgene.2021.100310
33. Das Laha, S., Dutta, S., Schäffner, A. R., & Das, M. (2020). Gene duplication and stress genomics in Brassicas: Current understanding and future prospects. *Journal of Plant Physiology, 255*, 153293. https://doi.org/10.1016/j.jplph.2020.153293
34. Basak, M., Chakraborty, S., Kundu, S., Dey, S., & Das, M. (2025). Identification, expression analyses of APETALA1 gene homologs in *Bambusa tulda* and heterologous validation of BtMADS14 in *Arabidopsis thaliana*. *Physiology and Molecular Biology of Plants*. https://doi.org/10.1007/s12298-025-01569-3
35. Pham, T. H., Tian, X., Zhao, H., Li, T., & Lu, L. (2024). Genome-wide characterization of COMT family and regulatory role of CsCOMT19 in melatonin synthesis in *Camellia sinensis*. *BMC Plant Biology, 24*(1), 51. https://doi.org/10.1186/s12870-024-04343-7
36. Vanholme, R., Ralph, J., Akiyama, T., Lu, F., Pazo, J. R., Kim, H., ... & Boerjan, W. (2010). Engineering traditional monolignols out of lignin by concomitant up-regulation of F5H1 and down-regulation of COMT in *Arabidopsis*. *The Plant Journal, 64*(6), 885–897. https://doi.org/10.1111/j.1365-313X.2010.04364.x

37. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., & Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology*, *139*(1), 5–17. https://doi.org/10.1104/pp.105.063743
38. Nicot, N., Hausman, J. F., Hoffmann, L., & Evers, D. (2005). Housekeeping gene selection for real-time RT-PCR normalization in tomato under abiotic stress. *Plant Molecular Biology Reporter*, *23*(4), 353–362. https://doi.org/10.1007/s11105-005-8887-9
39. Guénin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., & Gutierrez, L. (2009). Normalization of qRT-PCR data: The necessity of adopting a systematic, experimental conditions-specific, validation of references. *Biotechnology Letters*, *31*(4), 639–645. https://doi.org/10.1007/s10529-009-9990-1
40. Jain, M., Nijhawan, A., Tyagi, A. K., & Khurana, J. P. (2006). Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, *345*(2), 646–651. https://doi.org/10.1016/j.bbrc.2006.07.140
41. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, *3*(7), research0034. https://doi.org/10.1186/gb-2002-3-7-research0034
42. Saha, P., Das, S. S., & Chattopadhyay, S. (2017). Identification and expression profiling of COMT gene family in lignin biosynthesis of sorghum. *Plant Gene, 11*, 32–39. https://doi.org/10.1016/j.plgene.2017.06.002
43. Hu, R., Fan, C., Li, H., Zhang, Q., Fu, Y.-F. (2009). Selection of reference genes for real-time quantitative PCR normalization in tomato (*Solanum lycopersicum*) under abiotic stress conditions. *PLoS ONE*, *4*(11), e7703. https://doi.org/10.1371/journal.pone.0007703
44. Ye, J., Zhang, Y., Yu, G., Jiang, H., Zhang, D., Li, X., & Liu, Y. (2018). Comprehensive transcriptome analysis of lignin biosynthesis genes and transcription factors in moso bamboo (*Phyllostachys edulis*). *Plant Biotechnology Journal*, *17*(5), 905–917. https://doi.org/10.1111/pbi.12936
45. Zhang, Y., Huang, R., Pan, Y., Dai, L., Li, Y., Wang, Y., Zhang, C., & Zhang, Y. (2022). Identification and functional analysis of the caffeic acid O-methyltransferase

(COMT) gene family in rice (*Oryza sativa* L.). *Plants*, *11*(17), 2231. https://doi.org/10.3390/plants11172231

46. **2.** Sattler, S. E., Saballos, A., Xin, Z., Funnell-Harris, D. L., Pedersen, J. F., & Vermerris, W. (2009). Characterization of COMT gene expression in brown midrib mutants of sorghum. *Plant Physiology*, *150*(2), 584–595. https://doi.org/10.1104/pp.109.137653
47. Li, Y., Huang, X., Gao, J., Luo, X., Yang, Y., Liu, L., & Zhang, H. (2024). Selection and validation of reference genes in *Dendrocalamus brandisii* for quantitative real-time PCR. *Plants*, *13*(17), 2363. https://doi.org/10.3390/plants13172363
48. **4.** Fan, C., Ma, J., Guo, Q., Li, X., Wang, H., Lu, M., & Zhang, D. (2013). Selection of reference genes for quantitative real-time PCR in bamboo (*Phyllostachys edulis*). *PLoS ONE*, *8*(2), e56573. https://doi.org/10.1371/journal.pone.0056573
49. Gutierrez, L., Guénin, S., & Bellini, C. (2019). A general protocol for accurate gene expression analysis in plants. In D. Collinge (Ed.), *Plant Pathogenesis: Methods and Protocols* (Vol. 1933, pp. 1–17). Humana Press. https://doi.org/10.1007/978-1-4939-9040-0_1
50. **6.** Life Technologies. (2010). *Real-Time PCR handbook* (Rev. B). Thermo Fisher Scientific. Retrieved from https://www.gene-quantification.de/real-time-pcr-handbook-life-technologies-update-flr.pdf

Plagiarism report



turnitin Page 2 of 42 - Integrity Overview Submission ID trn:oid:::1:3264957182

7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Exclusions

▸ 2 Excluded Sources

Match Groups

45 Not Cited or Quoted 6%
Matches with neither in-text citation nor quotation marks

9 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
4% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Malay Das

Dr. Malay Das
Asst. Professor, Dept. of Life Sciences
Presidency University, Kolkata